数 学 小 品 译 丛

〔美〕R·洪斯贝格尔 著

[illegible] 译

数学瑰宝

第 三 辑

四川教育出版社

SHU XUE XIAO PIN YI CONG

数学小品译丛

数学瑰宝

（第三辑）

〔美〕 R·洪斯贝格尔 编

伍登祥 译

四川教育出版社

一九八九年 成都

特约编辑：胡师度

责任编辑：韩承训

封面设计：文小牛

版面设计：顾求实

数学瑰宝 （第三辑） 数学小品译丛

四川教育出版社出版发行 （成都盐道街三号）
四川省新华书店经销 绵阳新华印刷厂印刷

开本787×960毫米1/32 印张7.75 插页4 字数128千
1990年12月第一版 1990年12月第一次印刷
印数：1—1345册

ISBN7—5408—0200—6/G・198 定价：2.73元

作者为中译本写的附言

获悉《多尔恰尼丛书》中我的四卷《数学瑰宝》由译者译成中文，这实在是非常意外的事情。我为本书有了中译本殊感荣幸，欣然写出下面的简短附言，作为译本的开场白（不是用来代替本书原有的序言）。

译者希望把本书介绍给中国读者，我为此感到非常荣幸。相当初等的数学中有多少动人心弦的东西，这是值得注意的事情。这个译本里也寄托着我殷切的希望：愿读者将能揭示出数学中许多可喜的新东西。

洪斯贝格尔　1983.4.18

前言

美国数学协会的《多尔恰尼数学介绍性著作》丛书，是因机缘巧合而问世的。

纽约市立大学亨特学院多尔恰尼（Mary P. Dolciani）教授，本人是一位才华出众、工作热情的教师和作者，她一直在追求数学讲解工作中理想的上乘境界。

与此同时，协会已经获得本书的手稿，这是一卷短文集子，似乎并不完全适宜于收入协会现有的任何丛书，但是手稿内容引人入胜，文风清新，显然是值得出版的。

后来，水到渠成，多尔恰尼教授决定设立一笔周转资金来创办这套《数学介绍性著作》丛书，藉以实现她的目的。

本丛书的选题既要求清新的口语风格，也要求引人入胜的数学内容。预期，各卷都有丰富的习题，很多卷还附有解答。因此，这套丛书将能提供有价值的丰富材料，尤其是综述性课题的材料。我们设想，这套丛书是有数学才能的中学生可以看懂的，但水平更高的数学工作者也不能对此掉以轻心。

美国数学协会当然乐于接受多尔恰尼教授设置这套丛书的慷慨馈赠。她既是协会出版委员会的委员，又是管理委员会的委员，为协会服务，成绩斐然。管理委员会真诚愉快地决定将这套丛书冠以她的名字，以示敬意。

美国数学协会出版委员会主席
贝肯巴赫
（Edwin F. Beckenbach）

目　录

一、色图

H.L.多瓦特，D.T.芬克拜纳

1．导言 要是你喜欢打赌，而且想赢的话，那么只要至少有五人与你赌赌，你可以提出下面这个赌法。你只消这样提赌博条件，就是在聚赌的六人或更多的人中，要么（1）有三个人，其中的每个人都和其余二人曾经握过手，要么（2）有三个人，其中没有一个人曾经和另外两人握过手。

你也许会觉得这个赌法不大公平，无论你怎样设法引诱人家上圈套，你还是觉得你的五位赌友不至于蠢得来会接受这种赌法。事实上我们马上就要证明，你始终会是赢家；而且还要指出，每六人中必然总可以分成至少两个三人组，每一组都满足上述条件之一。

这个问题的另一种表述法于1953年出现在第十三届威廉·罗韦尔·普特南（William Lowell Putnam）大学生数学竞赛中①〔1〕，五年后作为问题E1321又登在美国数学月刊上〔2〕。自那以后发表了许多文章，以不同的方法拓广这个问

①有中译本，《普特南数学竞赛》，湖南科学技术出版社，1983年——译注。

题，简化已有的证明，探索这个问题与数学其他问题的关系。如果你喜欢解难题的话，读到这里就可以暂停一下，看看你是不是能解下面的问题。

E1321：

证明：任意六个人聚会时，其中有三个人要么相互认识，要么根本不认识。

2．再改变问题的提法，然后加以解决 上面的那种表述问题的方式一点儿也没有告诉我们如何用数学语言提问，但是一个问题一旦有了正确的数学表述，那多半也就等于是向问题的解决走了一大步了。稍加思索就可发现叙述不当之处。比如问题中提到了“人”便不恰当。用了“认识”，“握手”也不适当。重要的是有六个可以区分的对象，在其中的每两个对象之间，一个特别的关系要么成立要么不成立。这就告诉我们可以用几何的办法在平面上用六个点来代表六个人。如果两个人相识，我们就用一条线段（路径）把两点连结起来。如果他们不相识，那就不划线。经过变换之后问题就变成：

证明：所画出的图形要么含有一个三角形，要么含有这样的三点，这三点中任何一点都与其余两点不相连。

图中三点如果连了三条线，就叫满三角形；如果一条线也没连，就叫空三角形；如果只连了一条或两

条线则制文连角形

再稍微变化一下问题就更好解了。首先画出六点，其中任意三点都不共线。然后把每一点与其余各点用线段连结起来。如果两点所代表的人相识，就用蓝色线段连结这两点。如果两点所代表的人不相识，就用红线连结（也可以象图1所示那样用实线代蓝线，用虚线代红线）这样画出的图形叫完全双色图。令B表示双色图中蓝色三角形的个

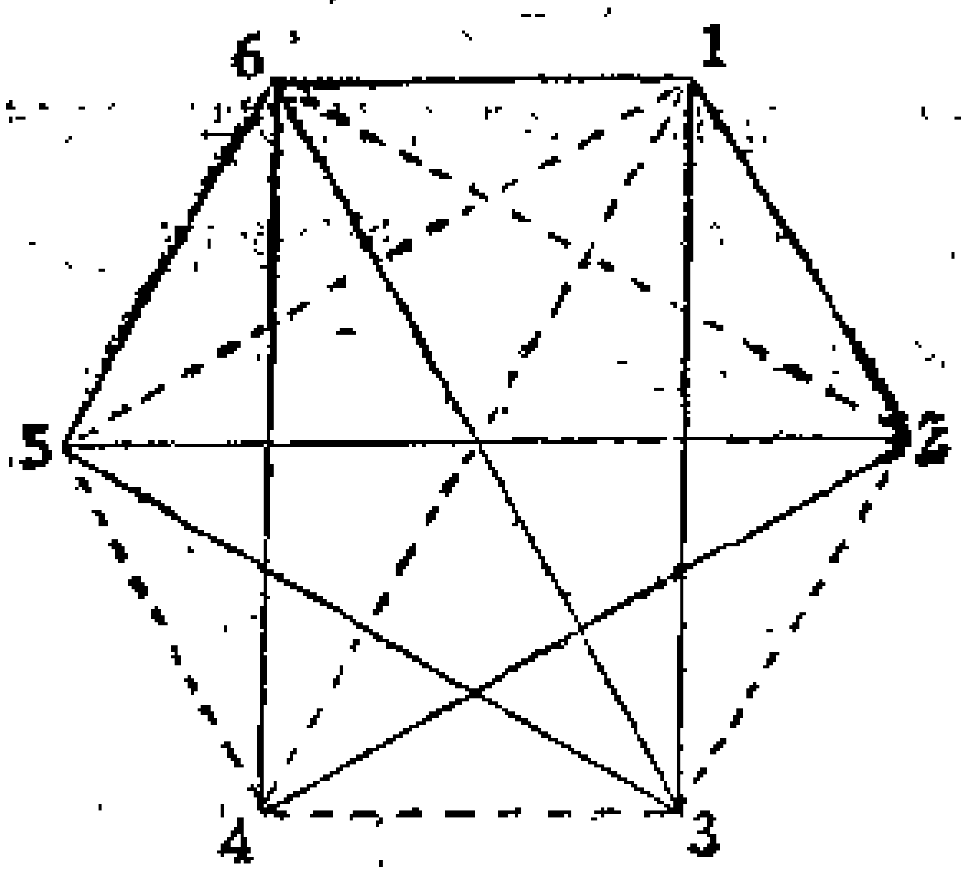


图1

数，R表示红色三角形的个数，M表示（混杂颜色的）欠三角形的个数。于是原问题变为：

证明：在任意的含六个顶点的双色图中

$$R+B\geqslant 1.$$

练习1：求图1的$R+B$的值。

问题提法改变以后，简单证明垂手可得。我们

来看任一顶点与把它和另外五个顶点相连的五条线。五条线中至少有三条线颜色相同，比如说都是蓝色。蓝线连的三点中如果任意两点间也是蓝线所连，则这两点与最初那一点便组成一个蓝色三角形；若三点中任何两点都不是蓝线所连，则这三点组成一个红三角形。因此$R+B\geqslant 1$。

其实我们最初这个问题要求的结论太弱了。我们后面将证明更强的结论：

$$R+B\geqslant 2.$$

这促使我们考虑进一步强化结论是否可能。但是仔细研究一下图2便知，（就本例而言，）$B=0$，$R=2$，所以$R+B=2$。

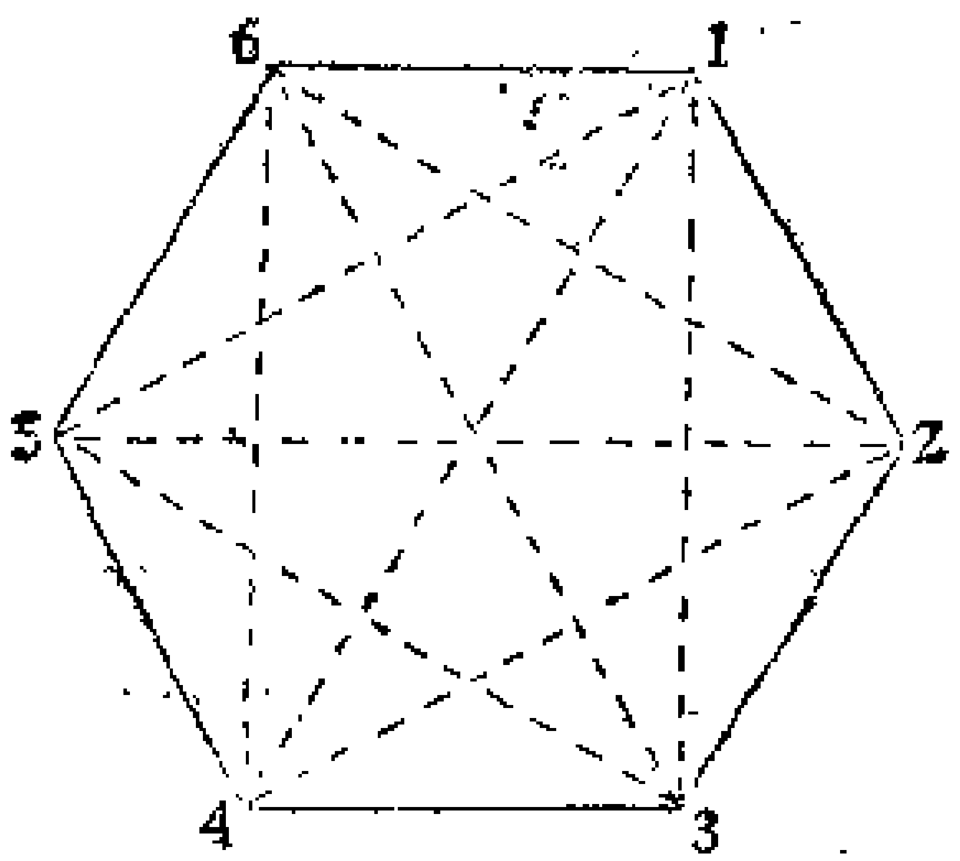


图2

关于完全双色图，人们还可以提许多问题。例如 完全双色图的红三角形最多有多少个？如果有

六个顶点，这种图中的各类三角形的总数是$C(6,3)=20$，即六个顶点中每次取三点的不同组合数。边（两顶点之间的线段）的总数是$C(6,2)=15$。如果所有的边都是红色，那么所有20个三角形都是红色。如果有14条边是红的，一条边是蓝的，那就只有16个红三角形，因为蓝边的两个顶点与剩下的4个顶点中的每一点都组成一个红蓝二色的三角形。如果有13条边是红的，2条是蓝的，那么就有12个或13个红三角形，这由2条蓝边相邻或不相邻来确定。依此类推。

练习2．如果一个完全双色图有6个顶点，12条红边，3条蓝边，那么$R+B$等于多少？

可以考虑更复杂的组合，不过我却来提一个与我们早先的问题有关的问题。我们把没有蓝色三角形的图叫无蓝图。试问，无蓝图的红三角形个数最少是多少？可以想象得到，限制B为零时$R+B$的最小值比对B不加限制时要大些。对6个顶点的图来说，图2证实了$B=0$时$\min(R+B)=2$。换句话说，在$B=0$的条件下$R+B$的极小值是不变的。对不足6个顶点的图有类似结果吗？或者对超过6个顶点的图有类似结果吗？我们对原问题的初步推广就是考虑含任意个顶点的图。

练习3。画一个与图2不同的无蓝双色图，要有6个顶点，并且恰好有两个红三角形。

3.N个(顶)点的双色图 我们考虑平面上$N\geqslant 3$个点，其中任意三点都不共线。完全可以把这些点当成是正N边形的顶点。将每两点用线段连结起来，组成一个N（顶）点完全图。线段（又叫棱边）总数为

$$C(N,2)=\frac{N(N-1)}{2}.$$

全部边或着红色或着蓝色，从而构成一个N点完全双色图。这个图中的三角形数目是

$$C(N,3)=\frac{N(N-1)(N-2)}{6}.$$

和原先一样，用B表示三边为蓝色的三角形数目，R表示三边为红色的三角形的数目，M代表杂色三角形的数目。于是

$$B+R+M=C(N,3).$$

我们先集中精力来看极图。极图就是对给定的N值，$R+B$取最小值的图。这个问题曾被A．W．固德曼〔3〕作为问题E1321的推广研究过。他证明了，极图的$R+B$值可用下列公式表示：

$$
\min(R+B)=\begin{cases} \dfrac{m(m-1)(m-2)}{3}, & N\text{为偶，} N=2m\text{时；} \\ \dfrac{2m(m-1)(4m+1)}{3}, & N\text{为奇，} N=4m+1\text{时；} \\ \dfrac{2m(m+1)(4m-1)}{3}, & N\text{为奇，} N=4m+3\text{时。} \end{cases}
$$

练习4、证明：六点无蓝双色图至少含有两个红色三角形。

为了对min $(R+B)$ 如何依赖于N有更深刻的感性认识，我们可以用上面的公式计算出下面这个表来。

表 1

N	min $(R+B)$
1～5	0
6	2
7	4
8	8
9	12
10	20
11	28
12	40
…	…

用原问题的语言来表达，我们可以这么说：九人相聚会，定有十二个或十二个以上不同的三人组，在这种三人组中要么三人互相认识，要么互不认识。

喜欢亲自动手试一试的人可以画出给定了顶点个数的极图来验证表中的某些数值。$N=1$，2，3的情形当然是太显然了，无需作图；$N=4$时也容易作出$R+B=0$的极图来。要发现$N=5$时的这种图就稍难些了。至于$N=6$的情形，我们已经提到过，图2就是一个$R+B=2$的极图，不过至今我们都还没有证明$R+B$恒大于1。

练习5. 验证表1中$N\leqslant 5$的各种情况。

4. 固德曼公式的一个证明 固德曼公式公布十三年后，A. J. 什温克〔4〕发表了一个更简单得多的证明，并用最大整数函数把结果表达得很简洁。对任意实数x，记号〔x〕表示不超过x的最大整数。例如$\left[\frac{3}{2}\right]=1$，〔2〕$=2$。什温克把固德曼公式写成

$$\min(R+B)=C(N,\ 3)-\left[\frac{N}{2}\left[\left(\frac{N-1}{2}\right)^2\right]\right],$$

上式中的两组方括号都表示最大整数函数。

什温克的证明分两部分。首先，他计算杂色三角形的个数，证明

$$M \leqslant \left[\frac{N}{2}\left[\left(\frac{N-1}{2}\right)^2\right]\right].$$

由于$R+B+M=C(N,\ 3)$，故

$$R+B \geq C(N,\ 3)-\left[\frac{N}{2}\left[\left(\frac{N-1}{2}\right)^2\right]\right].$$

证明的第二部分是对每个N指出一个极图，在此图上，上面最后一个式子中的等号关系成立。我们只给出什温克证明的第一部分，让你自己画达到极小值的图，画图很好玩的。

要使$R+B$极小我们就要让M取极大。因此，我们要确定N顶点图中杂色三角形的最大可能的个数。解这个问题的关键就是要明白在任意杂色三角形中一个顶点分隔同色边，两个顶点分隔异色边。如果在每个顶点我们计算颜色相反的关联边的对数，我们就正好把每个双色三角形计算了两次。令V_i，$i=1,\ 2,\ \cdots,\ N$，表示N个顶点，b_i表示与V_i关联的蓝边的条数，则与V_i关联的红线条数就是$N-1-b_i$，所以与V_i关联的异色边的对数是

$$b_i(N-1-b_i),$$

整个图中的双色三角形的总数是

$$M=\frac{1}{2}\sum_{i=1}^{N} b_i(N-1-b_i).$$

这是一个N项和，每项都是$b_1,\ \cdots,\ b_N$中的一个变量的平方。所以它的极大值不会超过N项的极大

值之和，而且函数

$$x(N-1-x)=\left(\frac{N-1}{2}\right)^2-\left(\frac{N-1}{2}-x\right)^2$$

在 $x=\frac{N-1}{2}$ 时取得极大值 $\left(\frac{N-1}{2}\right)^2$. 事实上由于和数$M$的每一项都取整数值，所以和的极大值就是 $\left[\left(\frac{N-1}{2}\right)^2\right]$. 最后，对全部$N$项求和，并注意$M$是取整数值的，我们就求得

$$\max M\leqslant\left[\frac{N}{2}\left[\left(\frac{N-1}{2}\right)^2\right]\right].$$

练习6. 证明什温克公式与 min $(R+B)$ 的固德曼公式是等价的。考虑四种情况 $N=4m+k$，式中 $k=0$，1，2，3.

练习7. 当 $N=2p$ 是偶数时，极图可以这样来画：把顶点分成各含p个点的两组，从一组的每点到另一组的每点画蓝边。用这样作出的图来验证表1在 $N=6$ 和 $N=8$ 时 min $(R+B)$ 的值。注意，在两种情形都有 $B=0$.

练习8. 当 $N=2p+1$ 是奇数时，表1中各项可以这样求出：标出N个点 P_1，P_2，…，P_p，Q_1，Q_2，…，Q_p，R. 对 $i=j\leqslant\frac{p}{2}$ 的全部i和j用蓝色把P_i与Q_j连结起来. $i\leqslant\frac{p}{2}$ 时，将P_i与R用蓝线连结；

同样地$j\leqslant\frac{p}{2}$时把Q_j与R用蓝色线连结起来。对$N=7$与$N=9$作这种图。注意，$N=7$时图满足$B=0$，但$N=9$时则是$B=2$。

练习9。请提供一般性的道理来证明，对于照练习8所说的办法作出的图，$B=\left[\frac{p}{2}\right]\left[\frac{p}{2}-1\right]$

5。极小无蓝图 完全双色图中单色三角形个数的一个下界既已知道，自然就要问极图中全部单色三角形是否可能属同种颜色。也就是说，如果已知$B=0$，我们想求N顶点图的$\min R$。R为极小的那种无蓝图称为极小无蓝图。任何无蓝图中红三角形的数目至少和表1中对给定N所列出的数目一样大，但是不知道对于某个N值极小无蓝图包含的单色三角形是不是必定比对于同一个N值的极图所包含的单色三角形数目多一些。

先检查N很小时的图我们便会发现，$N\leqslant8$时极图的确可能是无蓝图，但是若$N=9$，那末当允许有红三角形与蓝三角形时$R+B$的极小值便是12，而$B=0$时R的极小值就会是13了。

这个问题索夫〔5〕曾探讨过，他用P.厄尔迪什提供的道理把N为奇数的情形解决了。N为偶数的情形以前已为固德曼〔3〕所解决，他证明了只要N是偶数，就一定存在无蓝极图。索夫把

剩下的问题解决了．他证明了对每个奇数$N>7$，任一极图一定既含蓝三角形也含红三角形。更准确些说，当$N=4m+1$或$N=4m+3$，且$m>1$时，任何极图包含每种颜色的三角形至少$m(m-1)$个．而且一定存在一个极图，它恰好有$m(m-1)$个蓝色三角形．

自然还有一个问题应该解决，这就是求出顶点个数为奇数$N\geq 9$的无蓝极图中红三角形的极小数目．G·罗登〔6〕给了一个答案：若$N=2p+1$，$N\geq 11$，$B=0$，则

$$\min R=\frac{p(p-1)(2p-1)}{6}.$$

$N=9$，$B=0$时罗登证明了$\min R=13$，而在$N=9$的极图中$\min(R+B)=12$．

要是你一直在试验画双色图的具体例子，你自己可能已经发现了极图的几种一般画法．作为本节的结束，我们介绍一下罗登在〔6〕中得出我们未加证明所引用的那些结果所使用的几个方法．

我们从正N边形开始，依顺时针方向把顶点标成1到N．$N=4m$，$4m+1$或$4m+2$时，从每个顶点向它的两侧的m个最近的相邻点划蓝色边。其余所有顶点，一对对地都用红边连结起来．$m=1$，$N=6$时我们得到的就是图2，这是个极图，因为有两个红三角形，（1，3，5）和（2，4，

6），而没有蓝三角形。

$N=4m+3$时，仿上法，从每个顶点向它两侧最近的m个邻点划蓝线。只要$|i-j|=2m+1$，i到j也划一条蓝线。以$N=7$，$m=1$为例作出的是个极图，它有一个蓝色三角形（1，4，7）和三个红三角形（1，3，5），（2，4，6），（3，5，7），如图3所示。

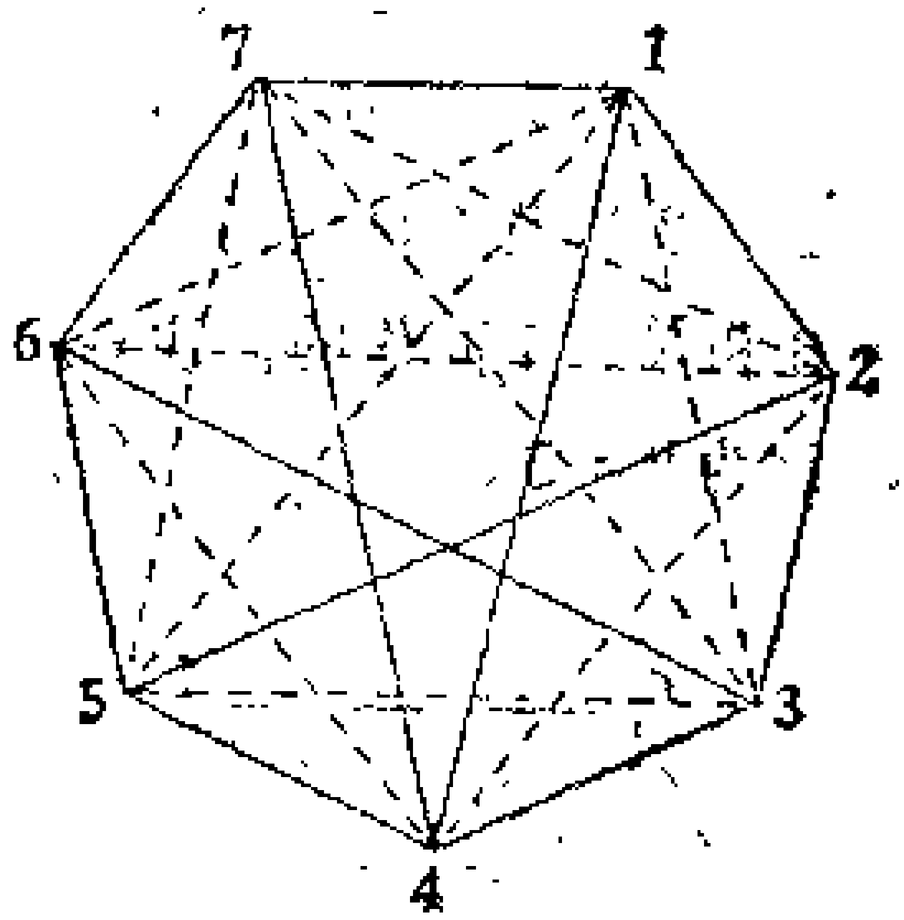


图 8

对每个N都能这样作出极图来。

练习10. 用罗登的作图法检验表1中$N=10$，11，12时$R+B$的值。求出每种情况下的R值和B值。

练习11. 造一个类似表1的表，对每个$N\leqslant 15$指出$\min(R+B)$的值和$B=0$时$\min R$的值。

有一种叫红网的画法可以给每个N值画出一个无蓝图。两点的编号之和为偶数时，在两点间划一条红线。换句话说，具有偶数编号的任意两个点用红线连结，具有奇数编号的任意两点也是用红线连结的。所以偶番号点构成一个红色网，奇番号点构成另一个红色网。由于任意三个正整数中至少有一对的和是偶数，所以每个三角形都有一条红色边，于是这一族的每个图都是无蓝图。N为偶数时，红网图总是极图，但N为奇数时红网图绝不会是极图（除去$N=1$，3这两个显然例外）。图4和图5便是为$N=6$和$N=7$而作的。注意，图5中的图不是极图，因为$R=5$。

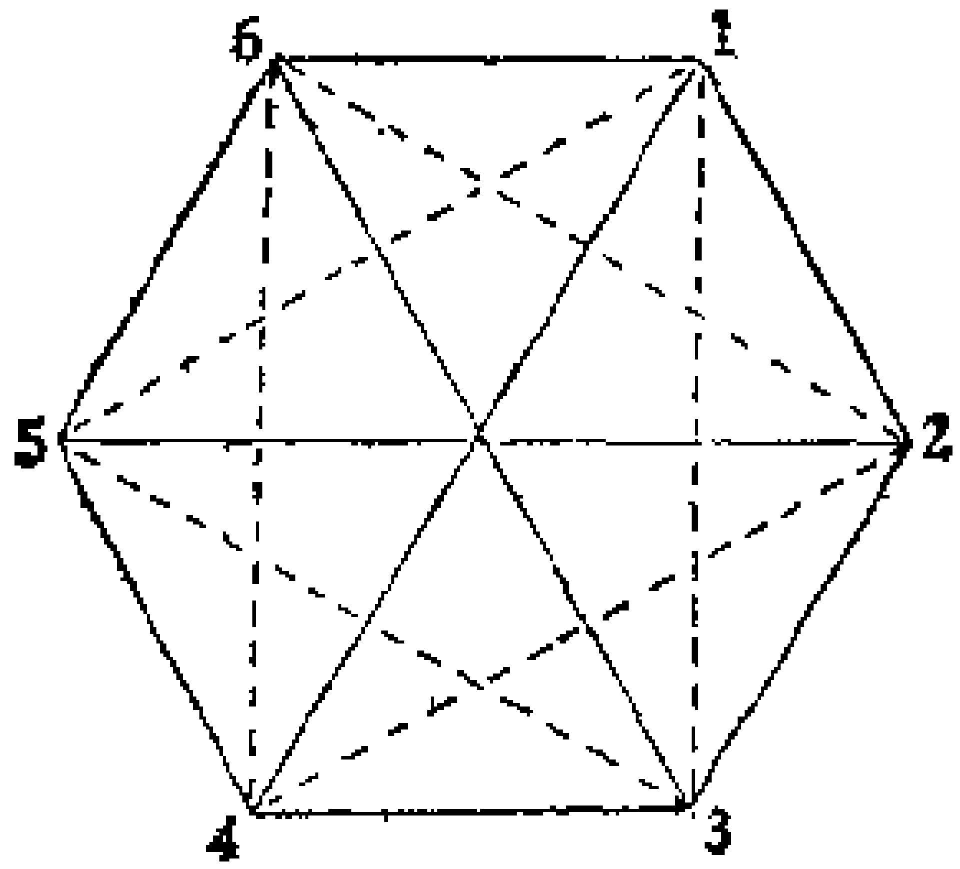


图 4

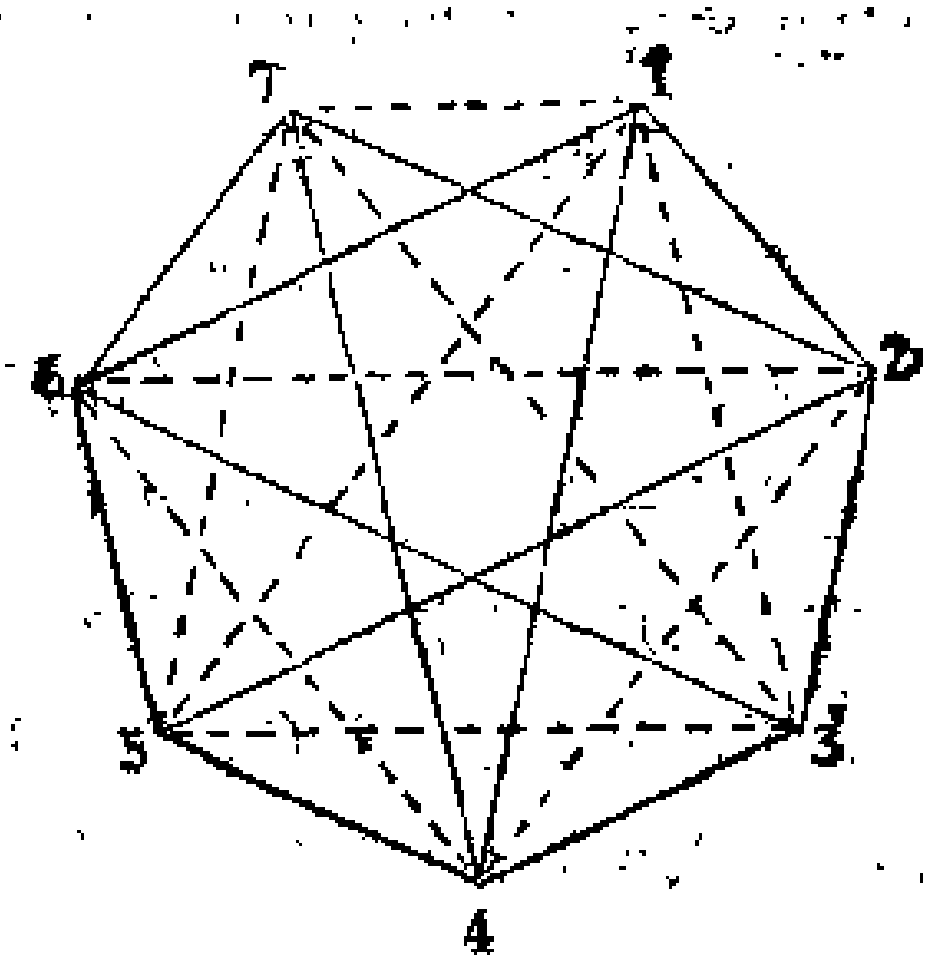


图 5

红网作图法是很有用的妙法，因为除了$N=5$，7，9这三个例外，对N的每个值用它都可以作出一个极小无蓝图来。

练习12．对$N=5$，7，9这几个例外情况，请各作一个无蓝极图。见〔5〕、〔6〕。

练习13．就$N=6$，8，11三种情形运用罗登的红网作图法；对每种情形检验R与B的值。

练习14．作为一个独立的（但并不太难的）问题请你求出N顶点完全单色图中不相交单色三角形的可能个数d_N。（不相交三角形无公共顶点。）特别地，请你证明d_N一定是$\left[\frac{N}{3}\right]$和$\left[\frac{N}{3}\right]-1$的中一个。见〔7〕。

6. 拉姆塞定理 我们的原始问题还可以有另外的推广。例如，我们可以研究要用两个以上的颜色来画边的完全图，或者不数三角形而数单色完全四边形或五边形。如何提出更一般的问题，使得原来的六人问题成为一般问题的特例？原始问题引导我们首先研究六顶点完全图，图中的棱边（顶点集的二元子集）用两种颜色来着色。我们计数的对象是单色三角形（顶点的三元子集，其每一个二元子集颜色都相同。）我们当初面临的挑战是要证明，6这个数已大得可以保证无论边是什么色我们数出的数绝不会是零。

为了说明这种拓广，我们来研究一下原始问题的一种特别扩展法。

（i）换掉六顶点完全图，让G表示平面上$N=100$个顶点的完全图。

（ii）不是给G的每个边一种颜色，而是给每个五边形（顶点的每个五元子集）一种颜色。共有$C(100, 5)$个这样的五边形，即多于7500万个。

（iii）不是只用两个色而是用四个色：蓝、红、绿、黄。要想把每个五边形区域真的涂上颜色，那当然是件太难完成的任务，因为五边形区域太多了，在交叠区域颜色一混合起来立刻就会把我们想显示的重要内容弄模糊。不过事实上我们只需

记住G的顶点的每个五元子集都涂上了四种颜色中的某一种就可以了。

(iv) 最后，给每种颜色配一个至少为5的整数。例如，给蓝色配上整数7。G的每个七边形（7个顶点）包含$C(7, 5)=21$个五边形子集，每个这种五边形都有四种颜色之一种。我们把7和蓝色联系起来时，所关心的是计G的七边形的个数，这些七边形的五边形子集必须是蓝色的。当然G的大多数七边形会包含不同色彩的五边形子集。如果某一七边形H的每个五边形子集正好都是蓝色的，我们就称H为蓝七边形。如果G的红八边形的个数（这里的红八边形是指它的$C(8, 5)=56$个五边形子集都是红色五边形的那种八边形），我们也可以指定整数8与红色相关联。现在假定我们关心的还有绿六边形和黄十边形。

给每种颜色都配好一个整数以后，问题就可以这样来表达：$N=100$是不是大得可以保证不管五边形是什么颜色，至少存在一个指定大小的单色图形。指定大小的单色图形指的是蓝七边形，红八边形，绿六边形或黄十边形。就本例而言，大得足以确保任意N顶点色图至少要有一个指定大小的单色图形的最小N值可表示为

$$R(7, 8, 6, 10; 5).$$

这个记号中有几个位置是写四个颜色的，在分号

前；另一位置在分号后。每个位置上的整数标明G的受人特别关心的子集的顶点个数，这些整数至少必须和最末位置上的那个整数一样大。最末位置上的那个数标明子集的大小，使得G的这么大的每个子集配了选定颜色中的一种色。本例最末数字是5，表示G的每个五边形子集配了一个色。分号前的那些数也表示G的子集的大小，每个颜色一个大小：蓝七边形，红八边形，绿六边形，黄十边形。

一般地，记号

$$R\ (k_1,\ \cdots k_t;\ r)\ ,$$

$$k_i \geq r \text{且} i = 1,\ \cdots,\ t$$

的含义如下：色图中的r元顶点子集配有t种颜色的一种，而我们关心的是存不存在一个K_i顶点组成的子集，使得这些顶点的r元子集配上颜色i。给顶点的r元子集每配上t种色的一种色，必定存在某个颜色i和某个K_i顶点子集，此子集的每个r元子集也具有所配那种色。为保证所说的情况发生，色图必须有的最小顶点数N就用记号$R(k_1,\ \cdots k_t;\ r)$表示。

从前面的讨论我们便知道$R\ (3,\ 3;\ 2) = 6$，因为在六或六个以上顶点的任意完全双色图中，必定存在一个用蓝边互连的三顶点集，或者一个用红边互连的三顶点集；而6便是使这一论断成立的最小数。

这个一般性问题的探讨肇源于英国人弗朗克·P·拉姆塞〔8〕1930年对形式逻辑所作的一项研究报告中。拉姆塞和坎特伯雷大主教是兄弟。数$R(k_1,\cdots k_t;r)$现在叫做拉姆塞数，断定拉姆塞数存在的论断就叫拉姆塞定理。在研究大量的组合问题，包括我们刚才一直在考虑的问题时，这个定理是一个有力的万用工具。例如问题E1321的解答便是拉姆塞定理结合$R(3,3;2)=6$而直接得到的。

拉姆塞定理的另一个特别情形就是直觉上很显然的盒子原理：如果把多于t个的物件放到t个盒子里去，必定有两个或多个物体占据同一个盒子。把t个盒子的每个盒子看成代表不同的颜色，于是这就和配色问题拉上关系了。每个物件（物件集合的一元子集）配上t种颜色的一种色。如果应配色的物件多于t个，那末至少有一对物件必定有相同的颜色。所以这时有$r=1$，$k_i=2$，$i=1$，…，t。由于$R(2,2,\cdots,2;1)=t+1$，所以拉姆塞定理就是盒子原理的一种表述。

拓广1953年普特南竞赛的双色图问题的早期文章中实际上就有一篇在讨论引向拉姆塞定理的一些推广问题。在这篇文章中R，E．格林武德和A．M格利森〔9〕保留了$r=2$，但考虑了t和k_1，…，k_t的各个特别的数值来探求拉姆塞数的性质，这些

我们列在表7中。

表　2

拉姆塞数$R(k_1, k_2; 2) = R(k_2, k_1; 2)$的值和界

k_1 \ k_2	2	3	4	5	6	7
2	2	3	4	5	6	7
3	3	6	9	14	≤19	
4	4	9	18	≤31		

$R(4, 4; 2) = 18$这个值表示的是，在任何18个或更多的人的人群中，必定有4个人互相认识，或者有4个人互相间根本不认识。但在17人或不足17人的人群中不一定就存在这种情况。同样$R(3, 5; 2) = 14$这个值指的就是在任意一个由14人组成的人群中必定有3个人互相认识，或者有5个人互助间根本不认识。$R(3, 5; 2)$值的确定可以先通过运用不等式（见〔10〕）

$$R(3,k;2) \leqslant \frac{k^2+3}{2}$$

于$k=5$的情形得出$R(3, 5; 2) \leqslant 14$，然后再来确立等式：考虑有13个顶点的这样一个双色图，对每个$i=1, \cdots, 13$，当且仅当$j=i\pm1$或$j=i\pm5$（取模13作运算）时V_i到V_j的边才着蓝色。这个图不含边全为蓝色的三角形，也不含边全为红色的完全五边形，所以$R(3, 5; 2) > 13$。

练习15. 用上述方法证明$R(3,5;2)=14$.

格林武德和格利森〔9〕还用数论和有限域论作为基本工具证明了

$$17\leqslant R(3,6;2)\leqslant 19,$$

$$R(3,3,3;2)=17,$$

$$42\leqslant R(3,3,3,3;2)\leqslant 66.$$

这三项结果的第一项在1968年时由J.E.格拉弗和J.雅克尔〔11〕作了改进，他们证明了$R(3,6;2)=18$，$R(3,7;2)=22$。另外的拉姆塞数$R(k_1,\cdots,k_t;2)$的性质就几乎不知道了。

$r=3$时，可以对拉姆塞数$R(k_1,\cdots,k_t;3)$作如下的解释：考虑N个顶点都在三维空间中的一个球面上的图。每一个三顶点集确定一个平面三角形区域，这个区域可以涂上t种颜色中的任意一种。于是$R(k_1,\cdots,k_t;3)$就是这样一个最小数，它使得如果$N\geq R(k_1,\cdots,k_t;3)$，则对某个$i$必定至少有一个$k_i$顶点的集，使这些顶点所确定的一切三角形区域都涂有第i种颜色。对于一般的r，我们可以考虑顶点位于r维空间中的球面上的一个N顶点图。每一个r顶点的集确定$(r-1)$维超平面的一个有限区域，它可以涂t种颜色中的任何一种颜色。如果N和对应的拉姆塞数一样大，那么某一个k_i顶点集就有这样一条性质：r个顶点在该集

中所确定的一切超平面区域涂有同样的颜色。

$r>2$时我们所知甚少，其中有一个结果就是J.G.卡尔布弗莱什〔12〕*得出的不等式

$$12\leqslant R(4,4;3)\leqslant 17.$$

写到这里我们就开始接触到一个组合问题的知识领域的前沿了。这个问题很容易，所以它的一个特例在二十多年前也就作为大学生攻读的问题提出过了，不过从此以后有关这个问题的信息增加得极少。如果你很幸运，发现了一个新的拉姆塞数，或者你更幸运的是发现了解决这类问题的新的有效方法，那么你的发现一定会使不少的数学家感兴趣。

在本文中我们回顾了1953年普特南竞赛中的一个貌似简单的问题与当前组合分析的数学研究中的某些方面的关系。作为普特南热持久不衰的进一步的证据，我们指出：加瑞·罗登积极地筹组了加里福尼亚工学院普特南队，他们在1971、1972、1973、1975和1976年度竞赛中得了头名，而格利森和格林武德在勒洛依·凯利的帮助下，把近25年的普南特竞赛试题和解答汇编成题集了。

---

*证明中补充：最近A·索布雷克证明了$14\leq(4,4,3)$，什温克证明$R(4,4,3)\leq 15$（见1978年1月美国数学会短评，A—29页。参见科学美国人1977年11月号马丁·加德纳关于拉姆塞数的一篇很有趣的文章）。

练习解答

1. 图1包含一个红三角形（1，4，5）和两个蓝三角形（1，3，6）和（3，5，6），所以$R+B=3$。

2. 应当考虑五种不同的情况：

a. 如果没有一个顶点属于两条蓝边，则$R=8$，$B=0$。

b. 如果有一个顶点属于两条蓝边，则$R=9$，$B=0$。

c. 如果有两个顶点属于两条蓝边，则$R=10$，$B=0$。

d. 如果有三个顶点属于两条蓝边，则$R=10$，$B=1$。

e. 如果有一个顶点属于三条蓝边，则$R=11$，$B=0$。

3. 图2与图4中的图都满足$R=2$，$B=0$。

4. 根据固德曼公式取$m=3$得$\min(R+B)=2$。如果$B=0$，则$R\geq 2$。

5. 就$N=4$的情形考虑一个正方形，其顶边和底边以及一条对角线都是蓝色。就$N=5$的情形，考虑一个正五边形，其边为蓝色，对角线为红色。

6. 考虑$N=4m$，$4m+1$，$4m+3$这三种情

况。在每种情况什温克公式都可以用代数法化为图德曼公式。

7. 最方便的是从圆周上等距分布的N个点开始，这N个点沿着圆周编号，所有号码是奇数的点为一个集，所有号码是偶数的点属于另一个集。

8. 照所说的办法画图，并注意$N=7$时$B=0$，$R=4$，而$N=9$时$B=2$，$R=10$。

9. 令t表示〔$p/2$〕。任意两个用P标记的顶点用红线连结，任意两个用Q标记的顶点也是用红线相连的。因此任何蓝三角形的构成必定是（P_i，R，Q_j）。但是如果$i>t$，R是用红线连到P_i的，如果$j>t$，R是用红线连到Q_j的，所以当且仅当$i\neq j$，$i\leqslant t$，$j\leqslant t$时（P_i，R，Q_j）是蓝的。对脚标i有t种选择，对于每个选定的i，脚标j有$t-1$种选择。所以$B=t(t-1)$，这就是要求的。

10. $N=10$时，$R=10$，$B=10$。

$N=11$时，$R=16$，$B=12$。

$N=12$时，$R=4$，$B=36$。

11.

N	$\min(R+B)$	$B=0$时$\min R$
1－5	0	0
6	2	2
7	4	4
8	8	8

9	12	13
10	20	20
11	28	30
12	40	40
13	52	55
14	70	70
15	88	91

12. $N=5$时利用边是蓝色、对角线是红色的五边形求出$R=0=B$。$N=7$时用練习8中所讲的作图法求出$R=4$，$B=0$。$N=9$时从圆周上循环编号的九个点开始，依次用红线连结下列三个点集：

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 1\},$$
$$\{1, 3, 5, 7, 4, 6, 8, 1\},$$
$$\{3, 9, 2, 8\},$$

于是得到$R=13$，$B=0$。

13. $N=6$时，$R=2$，$B=0$。

$N=8$时，$R=8$，$B=0$。

$N=11$时，$R=30$，$B=0$。

14. 每个三角形三个顶点，共N个顶点时，$d_N\leqslant[N/3]$。如果

$$d_N\leqslant\left[\frac{N}{3}\right]-2=\left[\frac{N-6}{3}\right],$$

则至少有六个顶点不在任何单色三角形中，这就与

问题E1321相抵触，所以

$$\left[\frac{N}{3}\right]-1\leqslant d_N\leqslant\left[\frac{N}{3}\right].$$

15. 根据已知不等式$R(3,\ 5;2)\leqslant\frac{5^2+3}{2}=14$。

正文中讲的有13个顶点的图中，当且仅当$|i-j|\in B$时顶点i与j才是蓝线连结的，并且

$$B=\{1,\ 5,\ 8,\ 12\},$$

当且仅当$|i-j|\in R$，且

$$R=\{2,\ 3,\ 4,\ 6,\ 7,\ 9,\ 10,\ 11\}$$

时i与j才是红线连结的。

注意，用红线或蓝线去连结一个顶点的模式时所有的顶点都是一样的。因此，如果有一个蓝三角形，就有一个含顶点1的蓝三角形。这种三角形的另外两个顶点的编号必定在集合$\{2,\ 6,\ 9,\ 13\}$中。但这些顶点的每一对都是用红线连结的，所以没有蓝三角形。

现在假定有一个含顶点1的完全红五边形P。由于1是用蓝线连到2，6，9，13的，所以P的另外4个顶点必定在$\{3,\ 4,\ 5,\ 7,\ 8,\ 10,\ 11,\ 12\}$中。

假定$3\in P$。由于3是经蓝线连到4，8，11的，所以只能把P选成$\{1,\ 3,\ 5,\ 7,\ 10\}$，但3与5是蓝线连结的，所以$3\notin P$。用同样的道理我

们可以依次证明12 $\notin P$，10 $\notin P$，4 $\notin P$。这样P就只有一种可能性，即｛1，5，7，8，11｝，其中7与8是蓝线连结的。所有含顶点1的任何完全五边形都不是红色的，根据图的对称性，对每个顶点来说这个断语都是对的。所以对这个图而言，$R+B=0$，这就证明了$R(3,5,2)=14$。

参考文献

1. L.E. Bush, The William Lowell Putnam Mathematical Competition, Amer. Math. Monthly, 60 (1953) 539—542.

2. H. Eves, Elementary problems and solutions, Amer. Math. Monthly, 65 (1958) 446.

3. A. W. Goodman, On sets of acquaintances and strangers at any party, Amer. Math. Monthly, 66 (1959) 778—783.

4. A. J. Schwenk, Acquaintance party problem, Amer. Math. Monthly, 79 (1972) 1113—1117.

5. L. Sauve, On chromatic graphs,

Amer. Math. Monthly, 68 (1961) 107—111.

6. G. Lorden, Blue-empty chromatic graphs, Amer. Math. Monthly, 69 (196?) 114—120.

7. J. W. Moon, Disjoint triangles in chromatic graphs, Math. Mag, 39 (1966) 259—261.

8. F. P. Ramsey, On a problem in formal logic, Proc. London Math Soc., Sec. Ser., 30 (1930) 264—286.

9. R. E. Greenwood and A. M. Gleason, Combinatorial relations and chromatic graphs, Canad. J. Math., 7 (1955) 1—7.

10. M. Behzad and G. Chartrand, Intoduction to the Theory of Graphs, Allyn and Bacon, Boston, 1971, 242—243.

11. J. E. Graver and J. Yackel, Some graph theoretic results associated with Ramsey's Theorem, J. Combinatorial Theory, 4 (1968) 125—175.

12. J. G. Kalbfleisch, On the Ra-

msey number N (4, 4, 3), Recent Progress in Combinatorics (W. T. Tutte, ed.), Academic Press, New York, 1969, 273—282.

二、公平分糕饼

K·瑞勃曼

每当两个人分到最后一块饼意见分歧时，一定有一个人会提议说“你切，我挑选”。这似乎是再公平不过了：切饼的人会切出两块他认为完全一样的饼来；挑选糕饼的人肯定也会觉得两半之中有一半饼的确是半个或比半个稍多。不过，在实际生活中对上述建议的回答通常是“不，不，你来切，我来选”，因为挑选的一方不消说是处于有利地位的。

挑选者得实利之处在于切饼者未必就能分得很准确，不过这倒不是理论困难之所在。我们只消假定这个数学饼可以用一把数学刀分割得很均匀。然而挑选者的另一个占便宜之处就不是那么容易排除的了。即使切割得万分准确，切饼者也只能保证自己得到他自认为恰好是一半的糕饼。选饼者对饼子的看法可能会很不相同，因此就有机会获得比他认为应得的一半还多的饼。

如果我们所追求的目标仅仅是两个人中每一人都得到自认为至少是半个糕饼的话，那么，“你先切

我后选”的分配原则已经相当满意了，尤其是在随机确定切饼者时。但是不止两人分饼子时，可以用什么公正的分法呢（如果有公正的分法的话）？如果参加分饼的人各怀不同的目的又怎么办？在回答这些问题时我们要考察几个分饼方案，其中还有几个方案具有令人惊讶的特性。

首先我们定下竞赛规则。假设有 n 个人参与分配一个数学饼（在整个讨论过程中，我们需要数学饼具有什么性质，它就具有什么性质）。任何两个人可以对饼子有不同看法，但是所有的人都必须是通情达理的。如果有人说这种话："那块不大，给我两块我才干。"或者"如果不从中间分给我，我就干脆什么也不要了。"那么这就叫做不通情达理。说得更精确点，通情达理这一点就要求：饼子未被动过或分成有限块，在任何人眼中饼子的任意部分都不失去或增加任何价值。还有，如果哪一块饼减少了，无论减少量是怎样微不足道，每个人都认为较小的饼块比起较大的饼块来说是整个饼的更少的一部分（当然，无须乎任何二人一致地认为究竟这两小部分是具体指的哪两部分）。当然大家都必须指定 1 为整块饼，而根本没有饼则指定为 0。

我们的目的就是把糕饼分成 n 块分给 n 个人，并满足下列判别准则：

C1：每个参加分饼的人都认为他所分得的一块，按他自己的尺度至少是整个饼的$1/n$；

事实上有各种各样的方案都可以达到这个目标，不过我们只讨论两个。第一个方案有点深奥，是"你先切我后挑"原则的直接推广。

我们姑且以三个人分饼为例来说明，三个分饼人是A_1（阿尔方斯），A_2（阿则伯特），A_3，（阿尔日隆）。随机地选其中之一，例如A_1，当切饼人。假定A_1(用他的数学刀)能把饼子切成三块：P_1，P_2，P_3；使得每一块对A_1来说代表了整个饼的三分之一。因为我们想要满足的是$C1$，所以分饼人A_2A_3只需说清楚三块中的哪一块他们觉得至少相当于三分之一；通情达理这一条又要求他们每个人认为至少有一块令他满意的。这一切统统可以表示在一个3×3矩阵中，矩阵的（i，j）元（即第i行第j列上的数）在A_i认为P_j块值得到至少三分之一时取值1，否则取值0。我们的假设保证这个矩阵的第一行有三个1，另外两行至少各有一个1。

比如，可能出现的就是下面这个矩阵：

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix},$$

这时三块饼可按下面方式分配：

$$\begin{pmatrix} 1 & 1 & \textcircled{1} \\ \textcircled{1} & 0 & 0 \\ 1 & \textcircled{1} & 0 \end{pmatrix},$$

即A_1得P_3这一块，A_2得P_1，A_3得P_2。

当然也可能矩阵像这样：

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

这就无法立刻得到令人满意的分配方案了。不过要是把P_1（或P_3）分给A_1，那么在A_2和A_3看来剩下的一定比三分之二还多，他们就会乐意用“你先切我后选”的原则来分剩下这部分。不难验证，在可能出现的49个矩阵中只有三个矩阵会引起这种额外的困难。但无论如何，三个人总是可以用满足判别准则C1的办法来分饼的。

讨论一般的n个人分饼问题时显然就要利用元为1或0的$n\times n$矩阵了。这种矩阵叫0-1矩阵，我们暂且顺便讨论一下它。在任意的$n\times n$矩阵中，（广义）对角线的定义是n个元的集合，其中任意两个元不在同一行或同一列，显然这样的对角线共有$n!$条。对于每条对角线，我们都可以把组成它的n个元作出一个乘积。这$n!$个乘积之和称为矩阵的恒定式。（注意，恒定式就是免去在一半的乘积前放负号的麻烦的行列式。）

对于0-1矩阵，当且仅当所有对角线上有0元时其恒定式才为0。有一条很有名的组合数学定理说，当且仅当矩阵含有许多0时每条对角线上才有0元。说精确点就是：

定理（弗洛贝纽斯-柯尼希） **$n\times n$ 0-1 矩阵的恒定式为0的充要条件，是该矩阵含有一个$r\times s$的0元子矩阵，这里$r+s=n+1$。**

比如刚才我们就看到了在3×3矩阵

$$\begin{pmatrix}1 & 1 & 1\\0 & 1 & 0\\0 & 1 & 0\end{pmatrix}$$

中每条对角线都含有一个0。处于（2，1），（2，3），（3，1），（3，3）位置上的各元构成一个2×2零元子矩阵。这条定理是制定n个人分饼方案的必须工具，这一方案是"你先切我后选"原则的推广。

n个人可以分饼而又满足判别准则C1，其证明方法是对n作归纳。我们已经证明二或三人办得到这件事，因此归纳法中的假设是：少于n的任意多个人分饼，都可以达到上述要求。

随意选择n个分饼人中的一个为切饼人，并假定他可能把饼切成n块，每一块对他来说都代表了饼的$1/n$。另外每个分饼人必须认定至少有一块是合乎要求的。于是这就可以用一个$n\times n$ 0-1矩阵来

表达，如果第i个人认为第j块饼代表了至少$\frac{1}{n}$个饼，则(i, j)元为1，否则为0。这个矩阵第一行中有n个1，其它每行中至少有一个1。

任何一条对角线，只要它的全部元都等于1，就是一个公平合理的分饼办法。所以只有在每条对角线都含一个0时才会出现困难。这时，弗洛贝纽斯-扣尼希定理保证一定存在一个由0组成的$r\times s$子矩阵，而$r+s=n+1$，因此我们就可以挑选一个r尽可能大的这种子矩阵。由于第一行只含1，所以$r\leqslant n-1$。

假定我们事实上能够找出这种子矩阵，其$r=n-1$。那么（如果必要的话，每列重新排列）矩阵就像这样（当$r+s=n+1$且$r=n-1$时，$s=2$）：

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & * & \cdots & * \end{pmatrix},$$

其中$*$是0或1。

第一个切块可以给切饼人，剩下的$n-1$个切块合在一起从剩下的$n-1$个分饼人的眼光来看比整块饼的$(n-1)/n$强。如果把这$n-1$个切块合在一起当成一份看待，那么根据归纳法的假设，这剩下的一份就可以由剩下的分饼人用满足C1的方法分享了！

末了我们假设，当我们找到了r取最大的0组成的$r\times s$子矩阵，我们有$r<n-1$。这时矩阵的样子是（必要时进行重新排列）：

$$
\begin{array}{c}
s-1\left\{\begin{array}{c} \\ \\ \end{array}\right. \\
r\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right.
\end{array}
\left(\begin{array}{cccc|c|cccc}
\multicolumn{4}{c}{\overbrace{\hphantom{1\quad 1\quad \cdots\quad 1}}^{s-1}} & & & & & \\
1 & 1 & \cdots & 1 & 1 & 1 & 1 & \cdots & 1 \\
\hline
 & & A & & * & & & * & \\
\hline
0 & 0 & \cdots & 0 & 0 & & & & \\
0 & 0 & \cdots & 0 & 0 & & & & \\
\vdots & \vdots & & \vdots & \vdots & & & * & \\
0 & 0 & \cdots & 0 & 0 & & & &
\end{array}\right)
$$

我们来看一下左上角的$(s-1)\times(s-1)$矩阵B中能找出的任意一个由0组成的$p\times q$子矩阵：

$$B=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ & & A & \end{pmatrix}.$$

B的这种$p\times q$子矩阵与0组成的$r\times s$子矩阵合并在一起就产生整个$n\times n$矩阵的一个由0组成的$(r+p)\times q$子矩阵。现在要是$(r+p)+q\geq n+1$的话，那就可以找到由0组成的一个子矩阵，其维数之和为$n+1$而其行数多于选定的$r\times s$子矩阵的行数。由于r选得最大，所以必定有：

$$(r+p)+q<n+1,$$

或者等价地表示成

$$p+q<n+1-r=s.$$

也就是说，$(s-1)\times(s-1)$ 矩阵B不含$p\times q$零元子矩阵，而$p+q=(s-1)+1$。于是再用一次弗洛贝纽斯—扣尼希定理得知并非B的一切对角线都能含零；B的任意一条无零元的对角线明确地提供了在头$s-1$个分饼人中分配头$s-1$个切块的一个满意的方法。剩下的r个人都认为这$s-1$个切块每块都值不到$1/n$，所以剩下的那些饼对这些人来说总价值是超过r/n的。我们的归纳法假设再次保证了这r个人可以用满足C 1 的办法来分享这些剩下的饼。

蕴藏在上面讨论中的方法是能够得出大家接受的分法的一个构造性方法。幸运的是我们的数学饼能够经受这项技巧所要求的一切切割与捏和处理，虽然平常的糕点如果经受这些处理早就成为碎屑了。但是对于不熟悉矩阵方法，归纳法和其它数学技巧的人，如果一旦遇上了必须在这些人中分配实际的饼子的困难局面，那么当他得知下面这个方案时一定会很高兴。

这个方法可以取名叫“动刀切后继续动”法，它可以用满足判别准则C 1 的方式把一个饼分给n个人。和第一个方法大不相同的是这里不用什么高深的数学，而且也不需要把饼子一而再再而三地切开又合拢。

饼子是矩形时就很容易直观地表现这个方法。刀的一端悬挂在空中，然后平稳缓慢地在饼面上移动。任何一位分饼人只要觉得切刀走完饼的$1/n$立刻便喊“停！”，于是切下这一块给他（若有札线，它可以随便割断）。由于没有别的分饼人比他更先喊停刀，所以他们一定都认为剩下那一份至少相当于整个饼的$(n-1)/n$。切刀继续地移动直到另外有人喊停为止，这时又给这个人切下一块。剩下的$n-2$个人放弃了两次先喊停刀的机会，他们肯定都认为剩下的一份至少值得到$(n-2)/n$。继续不断地这样切下去，一直到饼子切成n块为止，这时每个人都觉得他分到的一块至少值得到$1/n$。

我们很自然地会料到也许有些分饼人太贪心，偏要等到他觉得切刀可能切出比$1/n$多的一块饼时才喊“停刀”，不过想多分的企图却会冒少得的危险，因为不贪婪的人的行动自会保证他们自己每人得到$1/n$，因而不顾其他人的反应而天然地完成分割任务。

上述两个办法都完成了满足判别准则C1的任务，即：每个分饼人都认为他分得的那一块饼至少能值整个饼的$1/n$。但是这两个办法都无法消除这种抱怨：“我固然得的是$1/n$，但是他那一块比我这一块大些”。很显然这个人不仅满足于自己公平地分得的那一份，他还想和人家比。这就是说饼子

分成n块后，参加分饼的人给每一块赋予一个分数值，而不仅仅给予每个块0值（不接受）或1值（接受）。

要继续讨论下去我们就必须把通情达理的定义作一点修正。我们现在假设大家同意，饼子切成可数多个切块后，它的任何部份的价值不增也不减。这就要求我们把"一个饼"的概念略加拓广。我们不想探讨这个概念的一切推广，只提醒注意一点，就是某些饼块现在很可能十分难于拿出来上桌敬客。比如假设饼的上面挂了不同颜色的糖霜，尽管颜色不重叠，却可以绕来绕去绕出五花八门的图案来。具有给定的某种颜色的部分很可能处在互不连通的几块上，可能由一些环组成，而环心又是颜色各异的。现在的n块切饼可能是这样组成的，即让一块由某种特定颜色下的那部分饼构成。不管某一块饼原先是怎样杂乱无章地分布在饼的各处，对于这一块饼的价值每个人都必须作出前后一致、通情达理的评价。

整个饼子分成互不相交的n块后，n个分饼人的感觉如何可以简约地用$n\times n$矩阵表达出来，在这个矩阵$M=(m_{ij})$中，m_{ij}表示第i个分饼人认为是第j块所代表的那一部份饼子。注意，$0\leqslant m_{ij}\leqslant 1$，并且$M$的每一行之和必为1，即$\sum_j m_{ij}=1$。

把饼分为n块的每个分法产生一个$n\times n$矩阵，

考虑一切可能的分法得出的这种矩阵组成的集合，表示成Ω。关于Ω有一个定理，这个定理说出来容易但证起来却相当难。

定理（李雅普诺夫） 如果M_1，M_2，…，M_k是Ω中的矩阵，λ_1，λ_2，…λ_k是非负实数，满足$\lambda_1+\lambda_2+\cdots+\lambda_k=1$，则$\lambda_1M_1+\lambda_2M_2+\cdots+\lambda_kM_k$也在Ω中。

这个定理有一些明显结果可以用于我们的分饼问题。分饼法之一是第一块就取整个饼，另外的$n-1$块都是"空饼块"。通情达理性使我们由此分法得到矩阵：

$$M_1=\begin{pmatrix}1&0&0&\cdots&0\\1&0&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\1&0&0&\cdots&0\end{pmatrix}.$$

当然我们定全也可以取第二块为整个饼，这时有

$$M_2=\begin{pmatrix}0&1&0&\cdots&0\\0&1&0&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\0&1&0&\cdots&0\end{pmatrix}.$$

一般地，整个饼作为第k块而其它各块是空块的分法就得到矩阵M_k，其第k列由1组成而其它列全是0。

这就检证了n个矩阵M_1，M_2，…M_n的每一个都在Ω中，因而根据定理，矩阵

$$(1/n)M_1+(1/n)M_2+\cdots+(1/n)M_n$$

也在Ω中。即是说，一定有一个分割饼子的分法，这个分法产生下面的矩阵：

$$\begin{pmatrix} 1/n & 1/n & \cdots & 1/n \\ 1/n & 1/n & \cdots & 1/n \\ \vdots & \vdots & & \vdots \\ 1/n & 1/n & \cdots & 1/n \end{pmatrix}$$

这就验证了一个难以置信的论断：存在着将饼分割成n块在n个人中间进行分配且满足下面判别准则C2的分法。

$C2$：每个分得人都认为他分到了按他自己的标准是$1/n$的饼。

应该提醒一两句。首先，这纯粹是一个存在性的证明；如果你必须实际地去切割饼子，这些办法是丝毫帮不了忙的。其次，按此法分割出的饼块是乱七八糟的块块，所以只有用数学刀来切数学饼才可能切出这些块块来。

为了继续推演值得注意的特性，我们假设至少有一块饼使两个人意见分歧，把这块饼叫做P，而用Q表示剩下的饼。第一个人对P与Q的评估值表示成$m_1(P)$与$m_1(Q)$；第二个人对P、Q的评价是$m_2(P)$与$m_2(Q)$。显然必须有

$$m_1(P)+m_1(Q)=1,$$

$$m_2(P)+m_2(Q)=1;$$

我们假定了$m_1(P)\neq m_2(P)$，所以显然$m_1(Q)\neq m_2(Q)$。不失一般性，我们把两个人这样编号，使得$m_1(P)>m_2(P)$，因而$m_1(Q)<m_2(Q)$。还有我们也可以设$m_1(P)\geq m_2(Q)$。（如果情况不是这样的话，我们只消把这两位分饼人重新另外编号，交换P和Q就行）。心里记住这些假设然后就可以执行几项神秘的计算了，这些计算的重要性以后就会显露出来。

构作一个数：

$$\alpha=\frac{n\cdot m_1(P)}{n\cdot m_1(P)+1-m_2(Q)-m_1(P)}.$$

我们已经作过一些安排使得$m_1(P)\geq m_2(Q)>m_1(Q)$。因为$m_1(P)+m_1(Q)=1$，我们肯定有$m_1(P)>\frac{1}{2}\geq\frac{1}{n}$，所以得出结果：$n\cdot m_1(P)>1$。由于$m_1(P)$或$m_2(Q)$都不能超过1，所以$\alpha$的分母为正。此外我们还有$1-m_2(Q)-m_1(P)=m_2(P)-m_1(P)<0$，所以$\alpha$的分母比分子小。从而最后得出结论：$\alpha>1$，暂且不谈这与我们有何关系。

一个分饼法是令第一块为P，第二块为Q，其余全部饼块都是空块。这个具体分饼法产生矩阵：

$$M_1=\begin{pmatrix} m_1(P) & m_1(Q) & 0 & 0 & \cdots & 0\\ m_2(P) & m_2(Q) & 0 & 0 & \cdots & 0\\ * & * & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots\\ * & * & 0 & 0 & \cdots & 0\end{pmatrix}.$$

对于$k=2, 3, \cdots, n$，我们令M_k代表整饼为第k块的分法所产生的矩阵.因此跟从前一样，M_k的第k列上全是1，其它地方全是0. 由于$M_1, M_2, \cdots, M_n$都在Ω中，所以对于任何一组和为1的非负数组$\{\lambda_k\}$，$\lambda_1 M_1+\lambda_2 M_2+\cdots+\lambda_n M_n$也在$\Omega$中.

我们想把λ定义成下面的数：

$$\lambda_1=\frac{\alpha}{n\cdot m_1(P)},$$

$$\lambda_2=\frac{\alpha}{n\cdot m_1(P)}\left(m_1(P)-m_2(Q)\right),$$

$$\lambda_k=\frac{\alpha}{n} \qquad \text{当} k=3, 4, \cdots, n \text{时}.$$

由于我们已经假定了$m_1(P)\geq m_2(Q)$，所以我们有$\lambda_2\geq 0$；当$k=1, 3, 4, \cdots, n$时肯定$\lambda_k>0$. 而且这些数之和一定等于1，这一点我们马上就来验证：

$$\sum_k \lambda_K=\frac{\alpha}{n\cdot m_1(P)}+\frac{\alpha}{n\cdot m_1(P)}\left(m_1(P)-m_2(Q)\right)+(n-2)\frac{\alpha}{n}$$

$$=\alpha\left\{\frac{1+m_1(P)-m_2(Q)+(n-2)m_1(P)}{n\cdot m_1(P)}\right\}$$

$$=\alpha\left\{\frac{n\cdot m_1(P)+1-m_2(Q)-m_1(P)}{n\cdot m_1(P)}\right\}=1.$$

我们已经知道$M^*=\lambda_1 M_1+\lambda_2 M_2+\cdots+\lambda_n M_n$在

Ω中，若用m_{ij}^*来表示M^*的（i，j）元，则

$$m_{11}^* = \frac{\alpha}{n \cdot m_1(P)} m_1(P) = \frac{\alpha}{n},$$

$$m_{22}^* = \frac{\alpha}{n \cdot m_1(P)} m_2(Q) + \frac{\alpha}{n \cdot m_1(P)} (m_1(P) - m_2(Q)) = \frac{\alpha}{n},$$

$$m_{hh}^* = \frac{\alpha}{n} \qquad \text{当} h = 3, 4, \cdots, n \text{时}.$$

所以Ω含有一个矩阵，这个矩阵的对角线上的每个元都等于α/n。回想当初我们曾证明过$\alpha>1$。换句话说有一种把饼子分配给n个人的分法，满足下面著名的判别准则：

$C3$：每个分饼人都觉得，按照他自己的标准，他得到的那一份比$1/n$个饼还多。

不要忘记这个著名结果有一个假设条件作前提，那就是有两个人对某一切块看法有分歧。自然我们现在碰上的仍然是一个纯粹的存在定理，一点儿也没提到如何实现这个神妙的分法。令人惊奇的是，对意见分歧问题作出更强的假设，我们的确可以搞出一个满足准则$C3$的构造性的实施方案来分饼子。而且使用新方案时，饼子的每一切块实际上就是平常的条块。

我们的更强的假设条件是，根本没有饼子的那一条块会有两个人看法一致。如果只有两个人分饼，只消请他们各人独立地标出他自己认为的平分点。根据假设这两个标记不会重合。如果就在这两个标记之间的任何地点下刀切饼，那么两人中的任何一个都能得到他觉得比一半还多的饼。这一点在图 6 中解释得很清楚。

作为归纳法的假设，$n-1$个人中的每一个人都可以得到他认为比 $1/(n-1)$还多的一块(当然以我们更强的分歧条件为前提)。现在对n个人来说，叫每个分饼人在他认为能保证他得到$1/n$而愿意下刀之处作出标记。根据假定，各人看法都不一样，所

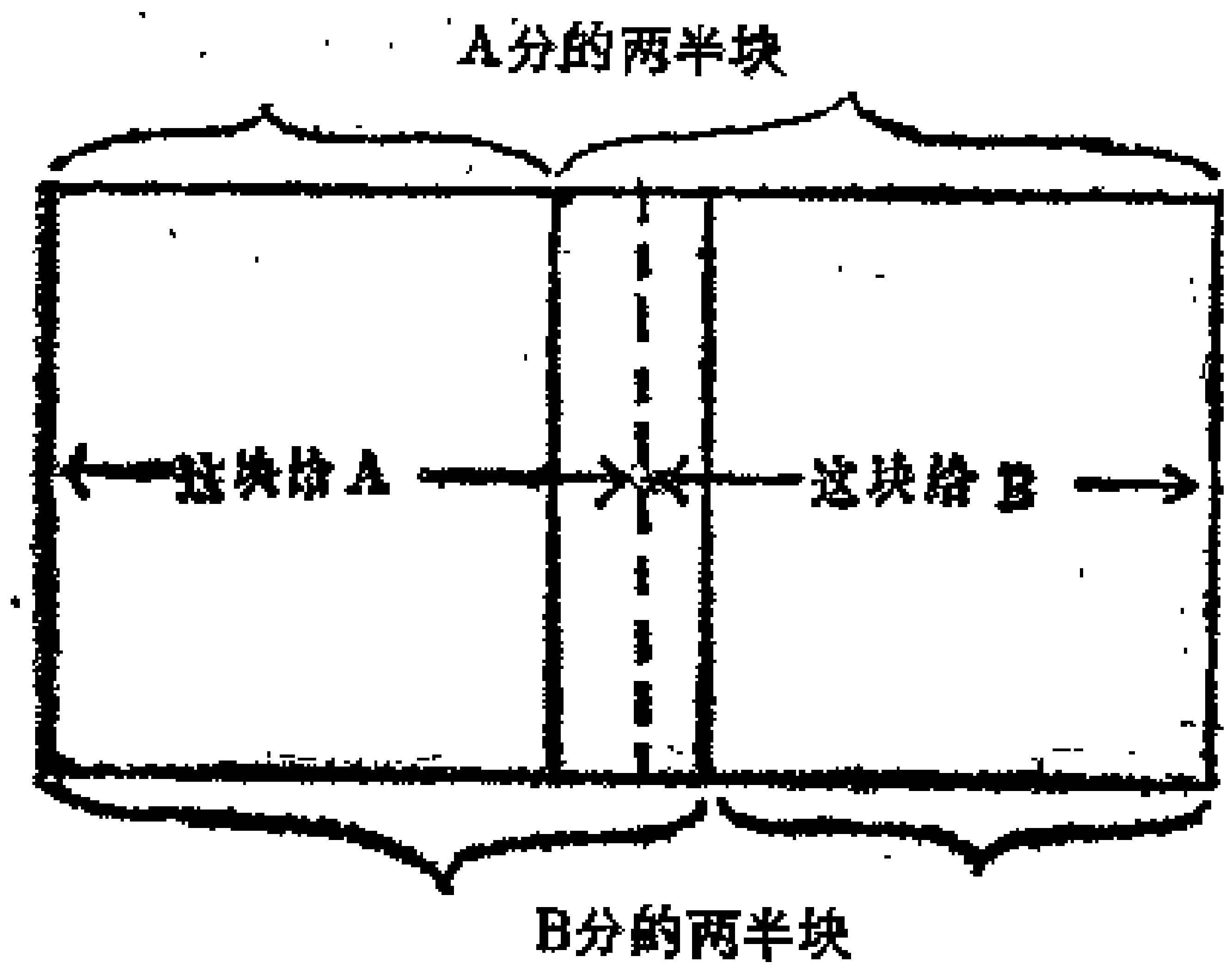


图 6

以没有标记重合，从左数起会有第一个和第二个标记。在这两个标记之间的任意点切一刀得到一块

饼，把它分给标出第一个标记的那个人；他会认为这比$1/n$多。另外的$n-1$个分饼人都会认为这一块比$1/n$少，于是他们可以如法炮制地把余下的饼分了，且满足判别准则$C3$。

这个结果证明，尽管分饼人之间意见分歧，还是可以找出人人认为公正的分配方法的。这个原则可以实用于分配不可分割的货物。下面我们不再在n个人中去分成份均匀的糕饼，而是考虑把k个不同的不可分割的物件进行分配的问题。这个问题的一个可能解决办法可以通过下面的例子解释清楚。

三个继承人要平分一份遗产：一座房子，一辆轿车，一艘小船，再加上一座小木屋。一个办法就是把这四件财产全都卖掉，再来平分所得的钱。但是三个继承人认为这样做不公正，因为他们对这几件东西值多少钱各人看法不同。所以他们说每个人应获得他所认为整笔财产应值价钱的三分之一，于是他们同意各自对遗产值钱多少作出估价来，这些资料汇总在下表中：

	#1	#2	#3
房子：	32000元	30000元	33000元
轿车：	3000元	2000元	4000元
小船：	3500元	4000元	2000元
小屋：	6500元	6000元	9000元
总价：	45000元	42000元	48000元
$\frac{1}{3}$份：	15000元	14000元	16000元

每件财产给予对该件财产估价最高的那个人。于是#3得到房子、轿车和小屋，按他的估价总共得到46000元；#2得到他觉得能值4000元的小船；#1一无所获。然而#1依法应得遗产15000元，而#2应得10000元。这两笔费应由#3支付，因为他得到的钱比他应得的公平份额还多30,000元。如果把额外这5000元拿来平分，那么每个继承人将比他认为的公平份额还多得了（实物或现金）1666.66元。

又是分东西的人意见有分歧这个前提使每个人都自以为得到的东西比平均份额多。我们再给一个例来说明这个现象。两个网球手在为每场比赛买新球时是平均分担费用的，以前是把用过了的球全部授给赢家了事，而今他们想出了更令人高兴的新花样来奖授用过的球。办法是：两人各自对要给的球出一个价；出价高的人得球，并支出一笔钱，这笔钱等于出价低的人的喊价加二人喊价之差之半。所以如果他们出的价分别为1.25元和0.75元，一个网球手支出1.00元得到他认为不止值那么多钱的球；另一位网球手得1.00元，而不要他认为值不到这笔钱的球。

如今我们已经明白了如何多得点，但愿我们今后永远得到至少是我们应得的那份吧。

鸣谢．我感谢H.W.库恩，L.E.杜宾斯，E.H.斯潘涅，因为本文的大部分材料都是取自他们三位的著作．我还感谢威廉·卢卡斯，是他引起我对这个题目注意的；感谢本册书的编辑在我准备手稿时给我的极宝贵的帮助．

注：人们不大可能确定得下来"你先切我后选"原则起源何处。这个原则在组合论上的推广归功于库恩〔3〕，尽管至人分饼的特例斯坦豪斯〔5〕早已提到过。

弗洛贝纽斯—扣尼希定理是组合论与图论中众多成果之一，它起源于偶图中关于匹配的扣尼希—埃哲伐利定理（或者同样可以说是来源于P·霍尔的著名"婚配定理"）。以这些思想为基础作出的一个推论可在〔4〕中找到。顺便提一句，无论是弗洛贝纽斯还是扣尼希，谁也可能不喜欢看到别人的名字和自己的大名并列。弗洛贝纽斯显然并未把扣尼希的成果看得很了不起；而扣尼希却觉得他自己的成果遭到了弗洛贝纽斯剽窃。这场争端是在扣尼希的措辞尖刻的脚注（〔2〕的第240页）中提到的。

"动刀切后继续动"原理与"最后减少者"原理是等价的，斯坦豪斯〔5〕说后一原理是S·巴纳赫和B·克纳斯脱发现的。

分饼人看待饼子必须通情达理，这条我们最初

所作的假定如果用数学语言来说的话就是要求每个人心中要有一个有限加性的、非原子的测度。过后我们又要求（非原子的）测度是可数加性的。这最后一点又要求全部饼块取自饼块的一个σ代数。

验证我们的判别准则$C2$和$C3$能够得到满足所使用的全部材料，都是杜宾斯和斯潘涅的论文〔1〕提供的。在这篇论文中证明了李雅普诺夫定理（实际上李雅普诺夫可能从来没有完全以这种形式叙述过），我们的应用例也可在这篇文章中找到。请注意，这个定理只说了Ω是凸集。

斯坦豪斯指出过两人要怎样分饼才可以都认为自己得到了比公平份额还多的饼，但我以前并未读到过推广到n个人的稍自然的推广。

分遗产的方案斯坦豪斯〔5〕说是克纳斯脱提出来的。用关于不等式的一项初等推理就可以证明，这个方案总是可以得出一笔非负的多余款项让他们去分。库恩〔3〕用线性规划中的对偶理论推导这个结果，他的推导优美极了。

网球分配方案萌发于我和E·基勒的闲谈中。

练习。（标＊的题比没有标记的常规题难。）

1．考虑下列矩阵：

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

找出0元组成的$r\times s$子矩阵，$r+s=7$，r为极大。如有必要重新安排行、列，使这个矩阵具有正文中的形状，然后在左上角的$(s-1)\times(s-1)$子矩阵中找出一条非零对角线来。

2．证明0—1矩阵的恒定式等于不含零的对角线的条数。

3．在切饼问题中假定分配第i个分饼人得k_i份饼，$k_1+k_2+\cdots+k_n=1$，$k_i\geq 0$。（正文中我们只考虑过对每个i，$k_i=1/n$这一特例。）试证明存在一种分饼的办法满足下面判别准则：对于每一对i，j，分饼人i认为分饼人j分得的饼正好值k_j。

*4．和题3一样，第i个分饼人的份额用k_i表示。证明有一种分饼的办法可以使第i位分饼人觉得他的那块饼值αk_i，这里$\alpha>1$。

5．每个参加分配的人都得到比他的份额多的一切分配方案中，分配的结局常常和他们互相间对各自的爱好的了解深度有关系。为了说明起见，假设在分遗产的例子中#1已经事先知道#2与#3对

四件财产的确切估价。你说，#1会如何喊价以确保他所得最多。对于#2与#3也请作同样的分析。如果三个人都事先知道了每个人的估价，你看他们会怎样采取行动？

6．就分遗产问题来看，请你证明，一般地讲总是有一份非负余额。

参考文献

1. L. E. Dubins and E. H. Spanier, How to cut a cake fairly, Amer. Math. Monthly, 68(1961) 1—17.

2. D. Kőnig, Theorie der endlichen und unendlichen Graphen, Leipzig, 1936.

3. H. W. Kuhn, "On games of fair division", Essays in Mathematical Economics in Honor of Oskar Morgenstern (M. Shubik, ed.), Princeton University Press, Princeton, N.J.1967.

4. L. Mirsky and H. Perfect, Systems of representatives, J. Math. Anal. Appl. 15 (1966) 520—568.

5. H. Steinhaus, The problem of

fair division, Econometrica, 16 (1948) 101—104.

三、一些非同寻常的整数序列

R·P·波亚士

1. 导论：调和级数 我们即将研究的序列的性质都和调和级数

$$(H) \qquad 1+\frac{1}{2}+\frac{1}{3}+\cdots$$

多多少少有些关系。有些人觉得很诧异，怎么(H)会是发散的，至少在刚接触时有这种感觉。这个级数一般说来是学生看到的第一个真正发散的级数（公项趋于0的发散级数）。对（H）的发散性的证明有几个很有启发性：

（1）（以10为基数的）一位数整数的倒数都比$\frac{1}{10}$大，它们一共是9个，所以它们在（H）中所占的份额大于$\frac{9}{10}$；二位数的整数的倒数都比$\frac{1}{100}$大，它们一共是90个，所以在（H）中它们占的份额也是大于$\frac{9}{10}$的。可以继续这样推论下去，所以我们在（H）中想找多大的部分和就可以找到多大的部分和。（想更详细的介绍，请读〔27〕的第98页。）

这个证明的更一般的形式在以2为基数时也一

样，不过没有这样明白地表达出来。这最早可以追溯到1350年左右尼古拉·奥热兹默的证明〔21〕，〔19〕；但它被后人忘记了，直到17世纪末才重新发现（H）是发散级数，而且证明也不同了。

（2） 这个证明是受了约翰·贝努利的证明（〔1〕，P.17）的启发而完成的；证明时假定我们学过正项收敛级数的代数知识，然后使用归谬法。假设(H)收敛于和数h，则

$$h=\sum_{n=1}^{\infty}\frac{1}{2n}+\sum_{n=1}^{\infty}\frac{1}{2n-1}=\frac{1}{2}h+\sum_{n=1}^{\infty}\frac{1}{2n-1},$$

于是

$$h-\frac{1}{2}h=\frac{1}{2}h=\frac{1}{2}+\frac{1}{4}+\cdots=1+\frac{1}{3}+\frac{1}{5}+\cdots.$$

上式中（最后一个等式）左边的每一项都小于右边的对应项，这就矛盾了。

（3） 贝努利自己的证明是这样的：还是假设（H）收敛于h，那么

$$h-1=\sum_{2}^{\infty}\frac{1}{n}=\sum_{2}^{1}\frac{n-1}{(n-1)n}$$

$$=\frac{1}{1\cdot2}+\frac{2}{2\cdot3}+\frac{3}{3\cdot4}+\cdots$$

$$=\frac{1}{1\cdot2}+\frac{1}{2\cdot3}+\frac{1}{3\cdot4}+\frac{1}{4\cdot5}+\cdots$$

$$+\frac{1}{2\cdot3}+\frac{1}{3\cdot4}+\frac{1}{4\cdot6}+\cdots$$

$$+\frac{1}{3\cdot 4}+\frac{1}{4\cdot 5}+\cdots$$

$$+\frac{1}{4\cdot 5}+\cdots.$$

即

$$h-1=\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)$$
$$+\left(\frac{1}{4}-\frac{1}{5}\right)+\cdots$$
$$+\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)$$
$$+\left(\frac{1}{4}-\frac{1}{5}\right)+\cdots$$
$$+\left(\frac{1}{3}-\frac{1}{4}\right)+\left(\frac{1}{4}-\frac{1}{5}\right)+\cdots$$
$$+\left(\frac{1}{4}-\frac{1}{5}\right)+\cdots$$
$$+\cdots.$$

显然第一行与第二行之和为1，第三行与第四行之和为$\frac{1}{2}$，第五行与第六行之和为$\frac{1}{3}$，等等，于是得

$$h-1=1+\frac{1}{2}+\frac{1}{3}+\cdots=h,$$

这是很显然的错误。这个式子从前叫“贝努利悖论”，它告诉我们，贝努利的同时代人是不理解贝努利的推理的。注意，这里要求作的事是得出一个错误的等式而不是要一个错误的不等式。

（4） 这个证明可能是最短的了，虽然（不像证明（1））要求我们知道收敛级数的性质。

设（H）收敛于h，则

$$h=\left(1+\frac{1}{2}\right)+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}\right)+\cdots$$

$$>2\cdot\frac{1}{2}+2\cdot\frac{1}{4}+2\cdot\frac{1}{6}+\cdots$$

$$=h.$$

这个证明似乎是很晚才发现的（曾几度被发现）。〔6〕；〔27〕，p.178；〔28〕。

（5） 人们常常依靠判断收敛或发散的积分检验法来判断（H）的发散性；如果精心地解说，我们会学到不少的知识，对以后很有好处。用几何方法来解释就很容易理解（见图7）。看图易知，曲线$y=\frac{1}{x}$下面从1到n的面积大于$\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}$，差额随$n$的增大而增大，即

$$\log n-\sum_{k=2}^{n}\frac{1}{k}$$

是递增的。（这里以及本文其它地方都有

$$\log n=\ln n=\int_1^n t^{-1}dt,$$

即，log表示以e为底的对数）。而且对数与和数之差还是有界的，因为从图中可以看出

$$\log n < 1 + \frac{1}{2} + \cdots + \frac{1}{n-1}.$$

所以

$$\log n - \sum_{k=2}^{n} \frac{1}{k}$$

趋于一个极限，因而

$$\sum_{k=2}^{n} \frac{1}{k} - \log n$$

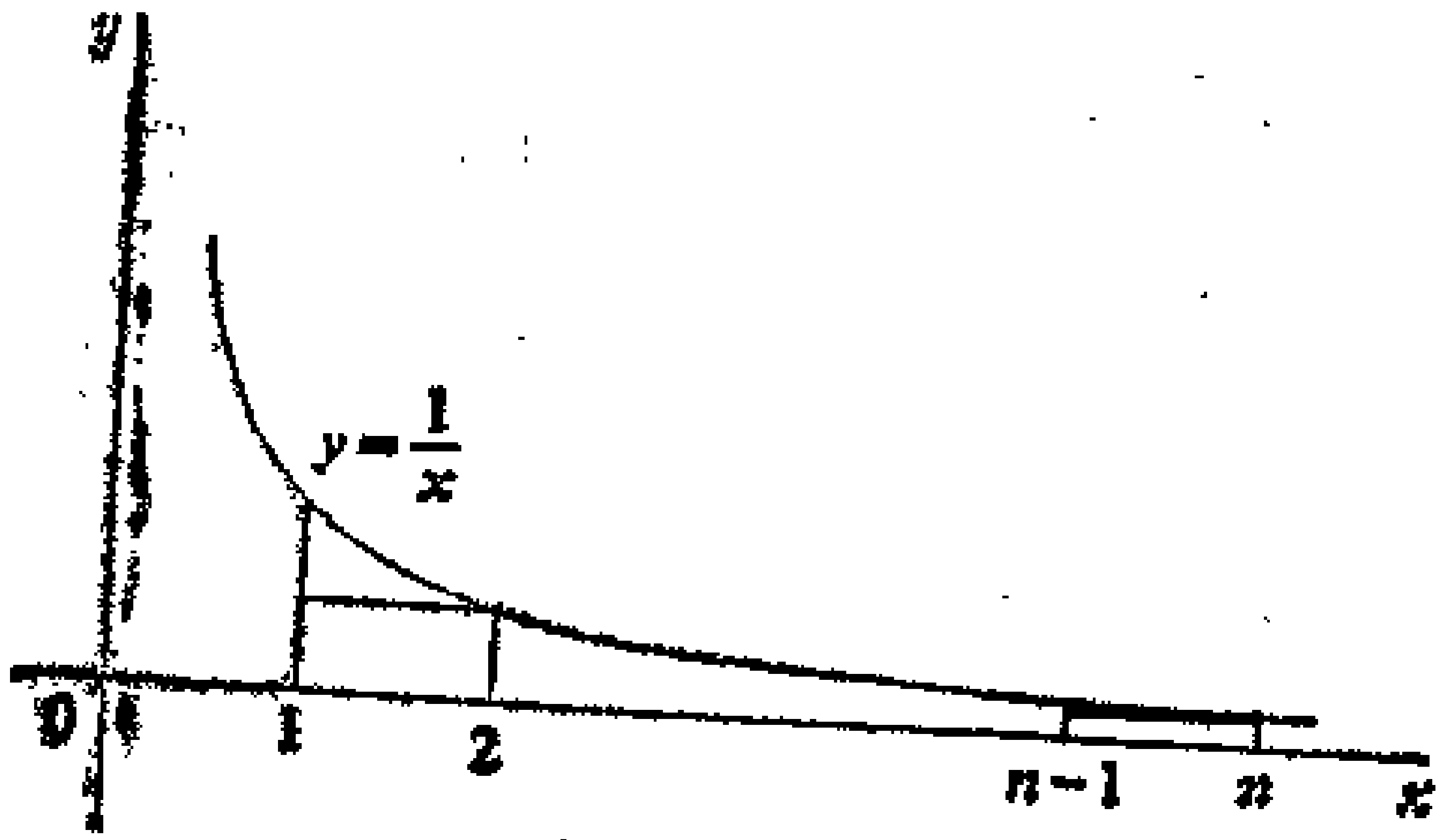


图 7

也趋向一个极限，这个极限用γ（有时用C）来表示，叫做欧拉常数。γ的十进制表示的头几位是0.5772156649，现在已经至少算出了20800位数字了〔2〕，但是我们还是不晓得γ究竟是个有理数呢还是无理数。如果γ实际上是一个有理数P/Q，那么我们现在就已经知道Q不止10000个十进位数字

〔2〕.

因为$n\to\infty$时$\log n\to\infty$，所以（H）是发散级数；它的第n个部分和很接近$\log n$。然而$\log n$是非常缓慢地趋向无穷（因为$2.30<\log 10<2.303$，所以$\log 10^{100}$约为230），所以我们可以说调和级数是十分缓慢地发散。事实上调和级数的部分和若要等于100，起码要用1.5×10^{43}个项才办得到（准确数目请查〔5〕和〔10〕p.167）。

当然还有发散得更慢的级数。看来不会有人去计算使$\sum\limits_{2}^{\infty}1/(n\log n)$的部分和超过10所必须有的准确项数（$10^{4321}$左右）；$\sum\limits_{3}^{\infty}1/(n\log n\ \log\log n)$的部分和要超过10所必须有的准确项数是一个$7\times10^{89}$位数，它比宇宙中的微粒数（目前认为约有$10^{80}$颗微粒）还大。

调和级数的发散性在娱乐消遣方面还有一些应用，最著名的一个问题就是：用完全一样的书（或砖）在桌边堆出一个悬臂拱，最上面那本书能伸出多远就伸出多远，请问这个悬臂拱（从理论上来讲）可以堆成多长？（〔23〕；〔17〕；〔12〕；〔8〕；〔10〕，p.167）。假定条件是完全理想的条件，用n本书能得到的伸出长度是调和级数的第n个部分和的一半，所以只要n充分地大，你是可

以要好长就可以得到好长的。（这个问题的另一个说法是在地板上堆书，伸出长度从最下面一本书算起，当然也就要多用一本书了。）由于$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}>2.08>2$，所以用4本书就可以得到一本书的伸出长度了。实际上看来得用5本书，因为真实的书既非刚性体又不是绝对粗糙的，如果我们太接近于最佳理论堆砌法，堆书时一定会滑。用31本书我们就可以得到两本书的伸出长度；用12367本书堆一个悬臂拱，伸出长度就有五本书那么长；若想悬臂拱的伸出长度为100本书长，那就得用4.05709 15001 19742 42417 27292 15083 27003 86982 29075 38568 62003 86928 96447 08306 50133 72179 599176×10^{86}本书！

上面的问题倒过来就是如图8那样用n根悬吊的细棒做成一个悬空的平衡活动架（这是R.C.布克提出来的）。这种车车灯式的活动架用喝饮料的麦管和线很容易就做得起。



图 8

2.稀疏的调和级数 在本节中我们要研究一个乍看起来令人惊讶的事实：如果我们从调和级数中取走用十进制表达时含零的整数的倒数，那末剩下的（子）级数是收敛级数．换句话说，我们从第一项起到$\frac{1}{9}$都保留，去掉$\frac{1}{10}$，保留$\frac{1}{11}$到$\frac{1}{19}$，去掉$\frac{1}{20}$，等等．

在前节中我曾经说过调和级数发散速度很慢；刚才提到的结果却使人觉得正好相反，调和级数发散得又有点快了，因为我们可以去掉许多项，依然还有一个发散级数．例如我们可以把全部复合数的倒数都去掉：如果p_n是第n个素数，则$\sum 1/p_n$是发散的．（这个结论在有关数论的书中有证明，例如〔15〕p.16.）很明显，从某个意义来说"大多数"整数是复合数．但是用十进制表示时"大多数"整数都含零．（以后我们简单一些说它们含零．）这个事实可能就不是那么明显的了．如果说这有点使人感到诧异的话，那只是因为我们大多数人对于大整数接触得太少了．事实上位数很多的大多数整数，用十进制表示时都含零，因为在$9\times10^{n-1}$个n位整数中不含零的整数只有9^n个．所以不含零的n位整数所占比例是$(9/10)^{n-1}$，这个数趋向零，但速度不太快，对于8位或更多位的数它才降到$\frac{1}{2}$以下．这就是说如果你打赌说随机地取一

个十进制中的n位整数，它都含有零；这只有$n\geq 8$你才会有一半的赌赢的可能性。这一事实表明，不含零的整数相对地稀少些，但这并未证明所谈到的级数是收敛的。

象上面那样稀疏了的调和级数有几个有趣的性质。它不仅收敛，而且收敛得极慢，慢得来逐项相加要求误差小于（比如说）2去求级数之和都不可能；不过用别的办法还是可以相当精确地计算它的和的。（有关讨论以及追溯到1914年的参考文献，请参看〔24〕和〔29〕。）

我们称此级数为（T）。我们用归并一些项在一起的办法来证明它收敛，先把1位数整数的倒数归在一起，然后把不含零的2位整数的倒数归在一起，继此以推。而1位整数的倒数都小于（或等于）1，它们一共有9个。不含零的2位整数的倒数都小于$\frac{1}{10}$，共计有9^2个；不含零的3位整数的倒数都小于$\frac{1}{100}$，共计有9^3个；一直这样推下去。因此将（T）与和数为90的几何级数

$$\frac{9}{1}+\frac{9^2}{10}+\frac{9^3}{10^2}+\cdots,$$

比较，得知（T）是收敛级数，而且还得到了(T)的和的上界。我们可以利用倒数表把前9项，前9^2

项，前9^3项，…等加起来，改进上界。但是趋近于零的速度很慢，十分粗糙的估计便可证实这点. n位或更少位的整数的所有倒数之后的余项包括9^{n+1}个（$n+1$）位数的倒数（每个都大于10^{-n-1}），9^{n+2}个（$n+2$）位数的倒数（每个都大于10^{-n-2}），等等。所以在$9+9^2+\cdots+9^n>9^n$项后余项大于

$$\left(\frac{9}{10}\right)^{n+1}+\left(\frac{9}{10}\right)^{n+2}+\cdots=10\times\left(\frac{9}{10}\right)^{n+1}.$$

说得更具体点，要把（T）中到13位数的整数的项加起来，我们一共得加$9^{13}>2.5\times10^{12}$以上的项才行。这时的余项才会大于$10\times\left(\frac{9}{10}\right)^{14}>2$。按最乐观的情况，设每项用1微秒的时间，我们也要用一个多月时间才计算得好。同样可以算出，即使在10^{18}项以后余项仍会比1大。如果我们真的把前66429项加起来，也就是说把用到5位或5位以下的整数的那些项加起来，得到的和大约是9.83，级数（T）显得来收敛得不错，然而在这种时候余项也比13还大，因为（T）的和算出来是23.103…，这里用的推算方法我们马上就要讲到。

很有意思的是，我们的确能算得这样准确。刚才我们已经看到，不能按常规途径去算，一直算到余项很小时为止。算余项时我们只须准确到足够保证和数的准确度要求就行了。在最方便之处，即与

n位整数对应的那些项之后，从上下两面来估计余项，就可以估计得足够准确。令S_n就表示n位整数的倒数和。例如考虑

$$S_2=\frac{1}{11}+\frac{1}{12}+\cdots+\frac{1}{99}.$$

S_2之后的余项首先就含有S_3的项，S_3的项中的1/111，1/112，…，1/119这9项都在1/120与1/110之间；1/121，1/122，…，1/129这9项都在1/130与1/120之间；…一直到1/991，1/999，这最末9项在1/1000与1/990之间。所以S_3的项在S_2以后的余项中所占的份额是介于

$$\frac{9}{10}\left(\frac{1}{12}+\frac{1}{13}+\cdots+\frac{1}{100}\right)$$

与

$$\frac{9}{10}\left(\frac{1}{11}+\frac{1}{12}+\cdots+\frac{1}{99}\right)$$

之间，换言之，是在$\frac{9}{10}U_2$与$\frac{9}{10}S_2$之间，这里的U_2是从S_2得来的，只不过把S_2的每个分母增加1罢了。同样地，我们加S_4项时实际上也是在$\left(\frac{9}{10}\right)^2U_2$与$\left(\frac{9}{10}\right)^2S_2$之间加；依此类推。最后，如果$R_2$是$S_2$后的余项，那么我们有

$$U_2\left(\frac{9}{10}+\left(\frac{9}{10}\right)^2+\cdots\right)$$

$$=9U_2<R_2<S_2\left(\frac{9}{10}+\left(\frac{9}{10}\right)^2+\cdots\right)=9S_2,$$

于是（T）的和T满足

$$S_1+S_2+9U_2<T$$
$$=S_1+S_2+R_2<S_1+10S_2。$$

使用倒数表（或袖珍计算器）很容易算出S_1，S_2和U_2来；最后的计算结果是$22.8027<T<23.4841$。

同样推理也可用到S_n上，不过准确度并无多大改善，即使把高速电子计算机动用起来也无济于事。R.拜意〔31〕用更精巧的估计法计算T，一直算到小数点后20位。准确到小数点后5位时，T＝23.10345。

完全类似地，我们可以把分母含0以外的某个数字的项都去掉。下表给出各个和数的准确到小数点后两位数字的值。（拜意曾把这些和计算到小数点后20位。）

删去的数字	和
1	16.18
2	19.26
3	20.57
4	21.33
5	21.83
6	22.21
7	22.49
8	22.73
9	22.92

我们还可以在以10以外的数为基数的进位制中来探讨这个问题〔24〕。特别是以100为基数时我们发现略去与某些整数对应的全部项后仍然得到一个收敛级数。这些整数的百进制表示式中不含百进制的某个具体的"数字"，即十进制中的某对具体的数字 ab，其中 $a\neq 0$；如果我们以1000为基数，略去某些整数的倒数也会得到一个收敛级数，这些整数的千进制表示式中不含100这个"数字"；更不必说如果我们略去的是十进制表示式中含两个或两个以上的零的所有整数的倒数了。

还应当提一提，不只是（T）收敛，因为不仅 $\sum 1/n$ 遍历不含零的整数时是收敛的，而且对于 α 略小于1的值，$\sum 1/n^{\alpha}$ 遍历不含零的整数时，也是收敛的，事实上 $\alpha>\log_{10}9=0.95\cdots$〔24〕。

级数（T）还可以从另外的角度来推广。我们不是去删除与含零整数对应的项，而是用一个小因子去乘每一项。假定某个整数含 h 个零，我们便用一个因子 $\varepsilon(h)\geq 0$ 去乘它的倒数。我们考虑级数

$$S=\sum_{n=1}^{\infty}\frac{\varepsilon(h)}{n}\qquad \varepsilon(0)=1.$$

我们有下述定理：

S随 $\sum_h \varepsilon(h)$ 收敛而收敛，随 $\sum_h \varepsilon(h)$ 发散而发散.

所有$h>0$时$\varepsilon(h)=0$的这种特殊情形，就是我们讨论过的（T）级数，而调和级数的发散则是对所有h均为$\varepsilon(h)=1$的情形。$\varepsilon(h)=1/h$的情形表明，即使把调和级数的每项都除以各项分母中零的个数也得不到收敛级数；如果我们晓得大整数k的位数大约不过$\log_{10}k$个，这也就不足为奇了。

为了证明这个定理，和以前一样，令S_n表示S的遍历n位整数的那个片段的和。S_n中出现的每个整数都在10^{n-1}与10^n之间。恰好有h个零的整数的数目是$9^{n-h}\binom{n-1}{h}$（因为首位不是零，所以这儿不是n而是$n-1$），它们在S中的份额在

$$\varepsilon(h)9^{n-h}\binom{n-1}{h}10^{-n}$$

的1倍与10倍之间。所以S_n也就在

$$10^{-n}\sum_{h=0}^{n-1}9^{n-h}\binom{n-1}{h}\varepsilon(h)$$

的1倍与10倍之间，于是S就在1乘下式与10乘下式之间：

$$\sum_{n=1}^{\infty}10^{-n}\sum_{h=0}^{n-1}9^{n-h}\binom{n-1}{h}\varepsilon(h)$$

$$=\sum_{h=0}^{\infty}\varepsilon(h)\sum_{n=h+1}^{\infty}9^{n-h}10^{-n}\binom{n-1}{h}$$

$$=\sum_{h=0}^{\infty}10^{-h}\varepsilon(h)\sum_{n=1}^{\infty}\left(\frac{9}{10}\right)^{n}\binom{n+h-1}{h}$$

$$=\sum_{h=0}^{\infty}10^{-h}\varepsilon(h)\frac{9}{10}\left(1-\frac{9}{10}\right)^{-h-1}$$

$$=9\sum_{h=0}^{\infty}\varepsilon(h)$$

所以S的收敛或发散是由$\sum\varepsilon(h)$的收敛或发散来确定的，这就是定理要求证的。

上面的定理还可以换一个角度去看〔25〕。第65页的级数S可以改写成$\sum\varepsilon(h)s_h$，这里的s_h是$\sum 1/i$，但i取遍所有恰好含h个零的整数。这就是要我们来研究序列$\{s_n\}$。〔25〕中已经证明这个序列是严格递减但却不趋于零（这和人们的直观感觉正好相反）。

首先我们证明 $s_0>s_1>s_2>\cdots$。

用$s(n,k)$表示只含n个零的k位整数的倒数和；这里$k\geq n+1$，因为所有整数的第一位数至少不会等于零。于是显然有

$s_n=s(n,n+1)+s(n,n+2)+s(n,n+3)+\cdots$.

组成$s(n+1,k+1)$的整数是具有$n+1$个零的$k+1$位整数。它要么等于10乘具有n个零的k位整

数；要么就是含$n+1$个零的k位整数，$k\geq n+2$，此整数右端末位是个非零数字。因而

$$s(n+1,k+1)<\frac{1}{10}s(n,k)+\frac{9}{10}s(n+1,k),$$

$$k\geq n+2.$$

而

$$s_{n+1}=s(n+1,n+2)+s(n+1,n+3)+\cdots,$$

再把前面的不等式用于上式第一项以后的每一项，我们得到

$$s_{n+1}<s(n+1,n+2)+\frac{1}{10}s(n,n+2)+\frac{9}{10}s(n+1,n+2)$$

$$+\frac{1}{10}s(n,n+3)+\frac{9}{10}s(n+1,n+3)$$

$$+\cdots$$

$$=s(n+1,n+2)+\frac{1}{10}\{s(n,n+2)+s(n,n+3)+\cdots\}$$

$$+\frac{9}{10}\{s(n+1,n+2)+s(n+1,n+3)$$

$$+\cdots\},$$

$$s_{n+1}<s(n+1,n+2)+\frac{1}{10}\{s_n-s(n,n+1)\}+\frac{9}{10}s_{n+1}.$$

而带$n+1$个零的$n+2$位数一定包含一个非零数字，后面接着是$n+1$个零，所以

$$s(n+1,n+2)=10^{-n-1}s(0,1),$$

$$s(n,n+1)=10^{-n}s(0,1).$$

最后得到

$$10^{-1}s_{n+1}<10^{-n-1}s(0,1)-10^{-n-1}s(0,1)+10^{-1}s_n$$

亦即

$$s_{n+1}<s_n$$

因为 $\{s_n\}$ 递减，故s_n趋向一个极限。现在证明这个极限是正的。虽然$\lim s_n$的数值至今尚未计算出来，但它比19.28大〔25〕。

s_n不趋于零，这是$\sum\varepsilon(h)$只要收敛则$\sum\varepsilon(h)s_h$就收敛的一个推论。设 $\lim\limits_{h\to\infty}s_h=0$，则对一切正数 k 我们可以找出整数 h_k，大到使 $s_{h_k}<1/k$。除h_k外对一切h都令$\varepsilon(h)=0$，并令$\varepsilon(h_k)=1/k$，$k=1,2,3,\cdots$。于是$\sum_h\varepsilon(h)=\sum\limits_{k=1}^{\infty}\varepsilon(h_k)=\sum\limits_{k=1}^{\infty}1/k$，这是要发散的调和级数。然而 $\sum\varepsilon(h)s_h=\sum\limits_{k=1}^{\infty}\varepsilon(h_k)s_{h_k}$ 逐项相比都小于 $\sum\limits_{k=1}^{\infty}1/k^2$。而后者是收敛的。所以$\sum\varepsilon(h)s_h$收敛。这和前面的定理是抵触的，因为在现在的情形依照定理 $\sum\varepsilon(h)s_h$发散。由于我们是从假设 $\lim\limits_{h\to\infty}s_h=0$ 为起点推论出矛盾来的，所以应该得出结论：s_h不趋于零。

3.数字符号在整数中的分布 如果你看到e，π或$\sqrt{2}$等无理数那一长串十进数字符号，你的印象多半是认为每个数字符号出现次数和另一个数字符

号的出现次数差不多是一样多，例如，可以说差不多有十分之一的数字符号是2。除了故意造的数字符号分布不均匀的数字而外，对于数字符号（即百进制中的单个数字符号）乃至对任何进位制中的数字符号你都会观察到同样的现象。要把这类论断说清楚说准确，需要用到几个专门术语。已给一个写成b进制表示的数字N，令d_n表示数字符号d在N的前n个数字中出现的次数。当$n\to\infty$时，如果$d_n/n\to 1/b$，并且对于一切基数b和我们可能挑选的任何数字符号，这一点也成立，我们便称数字N为正规数。实际上至今也还没有人证明过e，π或$\sqrt{2}$是正规数，但是的确有许许多多的正规数。更准确地说我们能够找得出一个区间序列，其总长可以任意地小，使得所有非正规数全都在其中；因此正规数比非正规数必定多得多。用测度论的行话来说就是非正规数集合的测度为零。这一事实通常更简明地说成：几乎所有实数都是正规数。关于证明，请看文献〔16〕或〔20〕。

从大多数数字都是正规数这一事实出发，可以推出一些令人惊骇的结果来。利用（比如说）ASCII代码赋予英语字母和标点符号一些数值（这在计算机上可以实现），可以把一篇英文文章转换成一个数字码。不妨假定我们把整套《不列颠百科全书》转换成了数字码，而且就把得出的这个数码当成一

个（很大的）整数B。B在基数为$B+1$（或更大）的计数制中是一个单独的数字符号，所以对于几乎一切实数（所有正规数），在其任何大于B为基的表示中，B都无穷多次地出现。我们如果把基选成10的某个幂，那么，说"数字符号"B在已给的数对该基的表示中出现无穷多次就等于是说百科全书的数字表示式在该数的普通十进制展开式中出现无穷多次。我们连制出现一次也没看到过，因为显示它出现一次所需要的十进位数的长度长得难以想象。顺便提醒一下，我们还不知道百科全书的数字表示式会在e、π、$\sqrt{2}$或类似数的十进制展开式中出现，因为还不知道有哪一个初等数学中的标准数是正规数。

找出明显表达出来的正规数的例子是很不容易的。这儿举出一例：0.123456789 10 11 12 13…，这个数的造法是把10进制下的相邻整数排成一行，前面放上一个小数点（〔20〕，p.112）。

我们不能期望整数的十进制表示是正规的 因为每个表示式中只有有限多个非零数字符号。但是你若仔细看一下那些用十进制写出的大整数（例如阶乘），你很可能又会得到数字符号分布得相当均匀的印象。大家也许会猜想对大多数整数情况恐怕是这样的，不过应对"大多数"给以某种解释。现在已经知道（见本文§2）十进表示中不含零的整

数是很稀少的，以致$\sum 1/n$对这些整数求和时是收敛的；然而在本文§1我们便已看到，$\sum 1/n$对全部整数求和时却是发散的。这就给我们提示了一种刻画一个整数集合S大小的方法。如果$\sum 1/n$在集合S上求和是收敛的，我们就说S"小"；不然就说S"大"。不含零的整数的集合就小。如果一个整数集合中的整数在其十进表示中（或b进表示法）时含零太少或太多，我们能不能断言这个整数集合一定小呢？由于我们的直觉总认为，以10为基数的表示法时，通常约有十分之一的数字符号是0，1，2，…，9中的某一个，所以我们需要有一个定义来表达某一数字符号在一个已知整数中所占的平均份量不是1/10这一说法。现在我们在任意的b进制下来探讨这个问题。

我们想给"n位数N中三分之一的数字符号是d"这一类话一个确定的意思。这句话的意思应当是，如果在N的n个数字符号中有k个是数字符号d，而$n=3k$；不过n不一定是3除得尽的。所以我们将这样来说：对于给定λ（$0<\lambda<1$），n位数N中数字符号d的比例是λ，或d在N中出现的频率是λ，如果d在N的b进制表示中的个数是$[\lambda N]$，即λN的整数部分。（如果我们使用更靠近λN的整数$[\lambda N+\frac{1}{2}]$，那就更符合我们的直觉；不过前一定义更便于处理，而这一点差别对结果没有什么影响。）

这个定义给d是否是以频率λ在N中出现这个问题以毫无歧义的回答，但是确定了N和d的数目并不能唯一地确定出λ。比如，$N=99^{40}$共有80位，其中11位都是8。由于$(1/7)\times 80=11.4$，我们可以说N的位数有1/7是8占据了的。我们也可以说N的位数有4/29或5/36是被8占据着的。在下面的叙述中，必须理解成先已选定λ。

我们打算证明在b进制的整数集合中，如果给定的数字符号以小于或大于$1/b$的频率出现，则这集合是小集合（在上面定义过的意义下）；如果某一给定数字符号以$1/b$的频率出现，则这集合就是大集合。这似乎支持了这样的思想：指定的一个数字符号在随机挑选的一个整数中通常出现"适当"的次数。而且$b\geq 4$时两个指定的数字符号通常是各以$1/b$的频率出现；然而若指定三个或三个以上数字符号，我们发现它们通常并不是各以$1/b$的频率出现。三个指定数字符号各以$1/b$频率出现的整数集合是小集合。例如，在十进制中，大多数整数都有十分之一的数字是0，另外十分之一的数字是1；但是0，1，2各占十分之一数字的整数就很少了。实际上大多数整数中数字符号是分布得很不整齐均匀的。

首先我们假设$0<\lambda<1/b$并指定一个数字符号d。令$m=[\lambda n]>0$，并考虑正好有m个数字

等于d的一个n位整数。如果我们不考虑左端第一位不为零这个条件，那么如果恰好有m位是d时就有

$$\binom{n}{m}=\frac{n!}{m!(n-m)!}$$

种方式去安排这些数字，有$(b-1)^{n-m}$种方式去安排剩下的数字。因此有m个或不到m个数字等于d时，n位整数的总个数就该是

$$\binom{n}{m}(b-1)^{n-m}+\binom{n}{m-1}(b-1)^{n-m+1}+\cdots+(b-1)^{n}. \quad (1)$$

然而首位数不能为数字0，(1)式就把至多有m个数字是d的n位整数的数目估计多了。由于我们想要证明对于d以频率λ出现的数N，$\sum 1/N$收敛，所以我们只需证明该级数对于（1）式所列举出的整数是收敛的。而这些整数的倒数都小于b^{-n+1}，所以倒数和也就不大于

$$\binom{n}{m}(b-1)^{n-m}b^{-n+1}+\binom{n}{m-1}(b-1)^{n-m+1}b^{-n+1}+\cdots+(b-1)^{n}b^{-n+1}.$$

如果记$p=1/b$，$q=(b-1)/b$，上式变成

$$b\left\{\binom{n}{m}p^{m}q^{n-m}+\binom{n}{m-1}p^{m-1}q^{n-m+1}+\cdots+q^{n}\right\}$$

$$=b\sum_{j=0}^{m}\binom{n}{j}p^{j}q^{n-j}. \quad (2)$$

可以看出，这个和数就是$(p+q)^n$按二项式定理展开的后$m+1$项。幸好这种和数都由概率论专家彻底研究过了：（2）式中的和数是成功概率为p失败概率为$q=1-p$的一项实验的n次独立试验（所谓贝努利试验）中有m或不足m次成功的概率。（参考关于概率的书，例如〔9〕，p.146以下各页）业已知道（参见〔9〕p.182起各页）$(p+q)^n$的展开式的最大项是m最接近np的那一项；一离开这个项往两端走各项都不断减小。由于$m=[\lambda n]$，$\lambda<1/b=p$，我们有$n\lambda<np$，于是（至少对于大n来说）最大项在（2）式的和数之外；所以（2）中的项从左到右渐渐减小，它们不足n个。因此(2)式的和不超过n乘第一项。所以我们得知n位整数在（1）所列的数的倒数和中所占比例小于

$$nb\binom{n}{m}p^m q^{n-m}=nb\frac{n!}{m!(n-m)!}p^m q^{n-m}. \qquad (3)$$

我们要想证明，让$n=1, 2, \cdots$对（3）求和会得到一个收敛级数。又很幸运，概率论专家已经有了答案：当n很大时，对于满足$0<\Lambda<1$的某个Λ，（3）式不超过Λ^n，所以我们真的得到一个收敛级数。

我们也可以用斯特灵公式来直接处理（3）式（例如参见〔18〕，p.531）。

b进制下的数字符号d在整数中以至多为λ的频

率出现，$\lambda<1/b$，这种整数的倒数和是收敛的我们现在已经证完了。而频率至少是$\mu>1/b$时呢，道理也一样，只不过用到的是二项展式的另一端罢了。

其次我们假设S是一个整数集合，在这些整数中d以$1/b$的比例出现；我们想要证明对于$k\in S$，$\sum 1/k$发散。这个结论从已经证明过的结论直接推不出来，因为我们要考虑的不是其数字符号d的频率介乎某个$\lambda<1/b$与某个$\mu>1/b$之间的数量的整数，而是其数字符号d的频率恰好是$1/b$的那些整数。

只消证明$\sum 1/k$在更小的集合S'上发散就够了，这个S'是S中的n可以被b除尽的那些n位整数k组成的集合。所以有$m=n/b$个数位上是d。由于我们需要的是S'中整数个数的下界，所以我们不必把这个数估大而是必须估计偏小些。$b>2$时我们就直接数含n/b个数字符号d的$(n-1)$位序列（可能从零开始数），因为在一个序列前任意放一个非0非d的b进制数字符号我们就得到一个S'中的整数。这种序列的个数是

$$\binom{n-1}{m}(b-1)^{n-m-1},$$

S'的每个数都小于b^n，所以S'中整数的倒数和大于

$$\binom{n-1}{m}(b-1)^{n-m-1}b^{-n}=b^{-1}\binom{n}{m}q^{n-m}p^m. \quad (4)$$

$b=2$时我们用稍稍不同的推理也得到一样的估值。最后一个式子是二项展开式的中央项的倍数，后者属$n^{\frac{1}{2}}$阶（或者可以利用斯特灵公式）。因而（4）式对n求和是发散的。

最后，当$b\geq 4$时我们考虑一些整数，在这些整数中数字符号d_1，d_2，$\cdots d_j$的出现频率为$1/b$。设d_i都不是0（要不然公式就必须修改）并记$m=[n/b]$。我们所考虑的n位整数的个数介于

$$\binom{n}{m}\binom{n-m}{m}\cdots\binom{n-m(j-1)}{m}(b-j)^{n-mj}$$

的两个倍数之间，而每个倒数则在b^{-n}与b^{-n+1}之间。若我们对一切n求和，则可得知全部倒数之和在

$$\sum_{n=1}^{\infty}\frac{n!}{(m!)^j(n-mj)!}\frac{(b-j)^{n-mj}}{b^n} \qquad (5)$$

的两个倍数之间。我在关于概率论的著作中没有发现（5）的渐近形式，但是利用斯特灵公式并进行一些代数演算我们就能够证明级数（5）的项原来都是在$n^{-j/2}$的两个常数倍数之间，所以$j>2$时（5）收敛，$j=1$或2时（5）发散。

4．第一位是1的整数　假定你有一组正整数，它们并不明显地偏向偶一边，也不明显地偏向奇一边，比如从电话号码簿中取出的许多街上的地址，

或者城市之间的里程；但不是超级市场里的杂货的价钱（它们偏向于以5或9结尾）。如果发觉到头来有一半左右的数是偶数，你会感到惊奇吗？我想，正好相反，这正是你预料之中的事，而且你还会竭力解释一番；有一半正整数都是偶数。现在假定我硬要你把最后这句话的意思说清楚，那么至少有两个合理的解决途径：实验法（“这是我们试了多次的结果”）和数学法（“n以前的偶整数的个数是$\frac{1}{2}n$或$\frac{1}{2}(n-1)$并且当$n\to\infty$时偶整数与n之前的全部整数之比趋于$\frac{1}{2}$”）人们常常喜欢用第二种，即数学的推演来说明实验观察的正确，哪怕数学并非就把实验模拟得很好。

这儿介绍几个术语是很有用的。设S是正整数的一个无穷集，在前n个正整数中有$S(n)$个是S的元素。如果$n\to\infty$时$S(n)/n\to\delta$，我们就说S的密度是δ，或者说δ是S的元素在整数中出现的相对频率，或者说δ是随机整数属于S的概率。能被10整除的整数集密度是1/10。平方数集合的密度是0；对应的实验观察是，在不明显偏向平方一边的大量的整数中通常只有相对说来极少几个平方数。集合的密度可以是正数，但集合仍可显得很不规则。例如可以证明（〔15〕，p.269）无平方因子数（相异素数之积）组成的集合密度是$6/\pi^2$，即约为0.6。你可以很有把握地打赌说，在一个随机的整数序列

中有一半以上的数都是无平方因子的。

并非一切集合S都有密度。你试想象有一个集合S，它由许多节相邻整数组成，节之间是长长的间隔，那么你自己就容易相信，并非所有的集合都一定有密度了。例如令S包含 2 与10（包括10）之间的整数，101与1000之间的整数，以此类推，即包含$10^{2n}+1$与10^{2n+1}之间的整数。那末

$$
\begin{aligned}
S(10^{2n}) &= S(10^{2n-1}) \\
&= (10-1)+(10^3-10^2)+\cdots+(10^{2n-1}-10^{2n-2}) \\
&= (10^{2n}-1)/11.
\end{aligned}
$$

所以当k取遍10的偶次幂时，$S(k)/k\to 1/11$，但k取遍10的奇次幂时却有$S(k)/k\to 10/11$。

我们刚才已经看到，象“随便一个整数有多大的可能性为 5 所除尽？”的这类问题有时有直观的答案，把问题所涉及到的集合的密度加以计算还可以使答案看起来很有道理。很多年以前有个人想起提这样一个问题：“正整数的第一位数是 1 的可能性有多大？”（假定这整数书写成通常的样子，从左往右念。）这个问题有意思的地方是，大多数人都猜答案是$\frac{1}{9}$，理由是有 9 个数字符号可以放在第一数位。然而这个猜测与实验大相径庭。事实上任何一个合理的随机整数表，比如杂货价格、道路里程或人口统计都表明 1 开头的整数多；的确，这类

表中三分之一的数几乎都是以1开头的。（这当然假定了表中无明显的偏倚，而不能象美国的航空里程表或大学新生年龄表那样。）要是第一位数字是1的所有整数构成的集合其密度算出来接近3/10的话，我们就会象用偶数集密度为$\frac{1}{2}$的数学结果去解释观察出的偶数频率一样地有权（或几乎有权）说"啊哈！所以我们观察到了我们所观察到的东西。"

然而不然，第一位是1的整数集合原来根本连密度也没有。事实上这个集具有我们为了构造无密度集所提出的那类结构。

我们来检查一下两个极端的子集吧：(a)具有n位或更少位的整数的集合；(b)集合(a)再加上$(n+1)$位整数的前10^n个整数。在(a)中我们有$9\times10^{k-1}$个k位整数，总共得$9+9\times10+\cdots+9\times10^{n-1}=10^n-1$个数，$k$位整数每组中有1/9的整数都以1开头。因而$m=10^n-1$时我们有$S(m)/m=1/9$。然而在($b$)中我们多加进了$10^n$个整数，它们都是1开头的。因为这时$m$等于$2\times10^n-1$了，我们有$S(m)/m=[(10^n-1)/9+10^n]\div(2\times10^n-1)$，其极限为5/9。因而$S$不可能有密度。有些人已经感觉到这里的理论和实验之间不一致了，应该有所解释，并且作了各种努力去解释。关于至今为止最详尽的讨论以及完整的文献，大家请读〔22〕，至于最新结果请看〔26〕与〔30〕。

解决我们的集合无密度这个数学困难的最简单方法是应当理解这一点：实在并没有什么了不起的理由非要我们把密度定义成$\lim S(n)/n$不可。我们可以把$S(n)/n$看成($\sum 1$；对于$k\le n$求和，$k\in S$)$\div$($\sum 1$，对于$k\le n$求和)。这使我们想到可以定义一个定义密度为

$$(\sum w(k), \text{遍历} k\le n \text{且} k\in S)\div(\sum w(k), \text{遍历} k\le n), \qquad (6)$$

式中$w(k)\ge 0$，并且$\sum_1^{\infty} w(k)$发散。大家将会看到，对$w(k)$只作一点宽松的约束，这就是相当合适的定义了。所谓“合适”我指的是如果S有普通意义下的密度，那么这广义密度就会和普通密度一样。如果$w(t)=1/t$，“宽松的约束”也就有了。即使在这个特殊情况，证明的技术情节也是够复杂的了，所以我不打算在这里给出证明。由于我们在§1中已经看到和数$\sum_{k=1}^{n} 1/k$几乎等于$\log n$，所以用$w(t)=1/t$的广义密度也是对数密度。

第一位数字是1的整数，其集合的对数密度是$\log_{10}2=0.30102999566\cdots$，这是一个令人瞩目的定理（在〔7〕中第一次给出明白的证明），现在我来描绘一下它的证明。上述结果可以看成是实验事实的数学模型，因为对数密度看来是和普通密度

简直一样好，如果不是更好的话。

这项结果的所有真正完备的证明都因技术细节太多而有些冗赘，但主导思想却相当简单。令S为第一位数字是1的整数的集合。S没有普通密度的证明意味着（而且这也容易验证）只检查一些特殊集合就够了：S的包含不足n位的一切整数的子集S_n。令L_i表示$\sum 1/k$，这里求和遍历以1开头的（$i+1$）位整数，即从10^i到$2\times 10^{i+1}$（包括第1个但不包括最后一个）的整数。我们从§1知道遍历从1到p的全部整数时，和数$\sum 1/k$（调和级数第p个部分和）几乎等于$\log p+\gamma$，此处γ是欧拉常数（p. 57）。那么，作为两个部分和之差的L_i几乎等于$\log(2\times 10^i)-\log(10^i)$，即，几乎等于$\log 2$。现令$L(n)$表示对$S$中不超过$n$的整数求和的$\sum 1/k$。对一切$i$求和得知$L(10^m-1)$，因而$L(10^m)$，几乎等于$m\log 2$。除以$\log 10^m=m\log 10$，我们得到$L(10^m)/m\log 10$，这几乎等于$\log 2/\log 10=\log_{10}2$。

随时要留意到所有的那些"几乎"虽然比较容易，然而却使人厌烦。不过我们可以简述一个替换的办法〔26〕，这个办法省去大量计算而且证明我们的集合S对于具有下列好性质的任何一个密度定义都有密度$\log_{10}2$；普通密度，自然缺这些性质的一条。令E和F为不相交整数集；令$2E$表示整数$2x$的集，这里$x\in E$；令$E+1$表示E往右移动1的集，

即整数$x+1$的集，这里$x\in E$。我们所需的性质是：若E有密度D则$E+1$也有密度D；若E有密度D则$2E$有密度$\frac{1}{2}D$；若E和F分别具有密度D_1和D_2，则其并集具有密度D_1+D_2。你们很容易就可以验证对数密度具有全部这些性质。

我们现在简要地说一下我们的证明，具有这些性质的任何密度都给我们的集合S赋予一个值$\log_{10}2$。稍加改变表明，若整数是用b进制表达的，赋予的值便是$\log_b 2$；若开头一个数位上是b进制中的q，则此值是$\log_b(1+1/q)$。

令$d(E)$为集E的密度，令变换T把E的每个整数x变成数对$2x, 2x+1$，使得E变成$2E$与$2E+1$的并集。另外，令$S(k)$表示（左边）由k开头的整数的集合，于是$S(1)$是S，$S(34)$是由34开头的整数的集合，以此类推。密度$d(S(k))$就简记成$d(k)$。

我们立刻就看出，把T用于S（1）就得出不相交集$S(2)$与$S(3)$的并集，所以根据密度d所假定的性质我们有$d(1)=d(2)+d(3)$。再用一次T我们有$d(1)=d(4)+d(5)+d(6)+d(7)$。继续用归纳法我们看出

$$d(1)=d(2^p)+d(2^p+1)+\cdots+d(2^{p+1}-1)。$$

把这个等式对$p=0,1,\cdots,m-1$写出再相加，得

$$md(1)=\sum_{i=1}^{2^m-1} d(i).$$

但是$\sum_{i=1}^{10-1}d(i)=1$,因为每个整数总有一个数字开头；同样$\sum_{i=10}^{10^2-1}d(i)=1$，如此等等。因此对10的直到$2^m$的每次幂，和数$\sum_{i=1}^{2^m-1}d(i)$计一个1，加上剩下部分所得的一个分数。所以和的整数部分是$[\log_{10}2^m]$，这与$[m\log_{10}2]$一样.但这也是$md(1)$的整数部分。由于对每个正整数m有$[md(1)]=[m\log_{10}2]$，所以$d(1)=\log_{10}2$必然成立。

练习

1.虽然很多地方都有堆书问题的公式(p.59)，但是常常没有推导。试推导之。推导时考虑从顶上数下来的第k本书上受的力，求对伸出端的力矩。

2.我们可以用稀疏的办法使发散级数收敛，当然也一样可以用浓缩的办法使收敛级数发散。假定k的十进制表示式含h个零，我们就用$\Delta(h)$乘$\sum 1/k^2$的对应项.试证$\Delta(h)=90^h$时得到的级数依然收敛；而当$\Delta(h)=91^h$时则发散.(Amer.Math.Monthly,

84(1977)652, problem E6275; 86(1979)58).

3.试估计其十进制表示式只含奇数字的整数的倒数和.(School Science and Mathematics, April 1975, problem 3555.)

4.试估计其十进制数字全不相同的整数的倒数和。(Amer.Math.Monthly, 82 (1975) 401, problem E2533; 83 (1976) 570.)

5.试证有无穷多个素数在其十进表示式中都含零。

6.在§2中其倒数被删去的那些整数有密度吗?是对数密度吗?

7.令S表示整数集合，这些整数的十进制数字(从右到左)构成不减序列，例如，766654430.试证在S上求和时$\sum 1/k$是收敛的，并估计这个和的大小。

参考文献

1. Jakob [Jacques, James] Bernoulli, Uber unendliche Reihen (1689—1704) (G.Kowalewski, ed. and trans.), Leipzig, 1909.

2. R.P.Brent, Computation of the regular continued fraction for Euler's co-

nstant, Math. of Comp, 31 (1977) 771—777.

3. R. P. Boas, Jr, A Primer of Real Functions, 2nd ed, Math. Assoc. of America, Washington, D.C, 1972.

4. ——, Partial sums of infinite series, and how they grow, Amer. Math. Monthly, 84 (1977) 237—258.

5. ——, and J. W. Wrench, Jr, Partial sums of the harmonic series, Amer. Math. Monthly, 78 (1971) 864—870.

6. M. R. Chowdhury, The harmonic series again, Math. Gaz, 59 (1975) 186.

7. R. L. Duncan, Note on the initial digit problem, Fibonacci Quart, 7(1969) 474—475.

8. L. Eisner, Leaning tower of The Physical Review, Amer. J. Physics, 27 (1959) 121.

9. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed, Wiley, New York, 1968.

10. M. Gardner, Martin Gardner's Sixth Book of Mathematical Games from

Scientific American, Scribner's, New York, 1971.

11.——,Mathematical games, Scientific American, March and April, 1975.

12.G.Gamow, Matter, Earth and Sky, Prentice-Hall, Englewood Cliffs, N.J, 1958.

13.H.Halberstam and K.F.Roth, Sequences, vol. 1,Clarendon press, Oxford, 1966.

14.G.H.Hardy,Divergent Series,Clarendon press, Oxford, 1949.

15.——,and E.M.Wright,An Introduction to the Theory of Numbers, 3rd ed, Clarendon Press, Oxford, 1954.

16. M. Kac, Statistical Independence in Probability, Analysis and Number Theory,Math. Assoc. of America,Washington, D.C, 1959.

17. P.B.Johnson, Leaning tower of lire, Amer. J. Physics, 23 (1955) 240.

18.K.Knopp,Theory and Application of Infinite Series, Blackie, London and Glasgow, 1928.

19. J.E.Murdoch, Review of [21], Scripta Math, 27 (1964) 67—91.

20. I.Niven, Irrational Numbers, Math. Assoc. of America, Washington, D.C, 1956.

21. N.Oresme, Quaestiones super geometriam Euclidis (H.L.L.Busard, ed), Janus, supplements, vol.3, Leiden, 1961.

22. R.A.Raimi, The first digit problem, Amer. Math. Monthly, 83 (1976) 521—538.

23. R.T.Sharp, problem 52, Pi Mu Epsilon J, 1 (1953) 322; (1954) 411.

24. A.D.Wadhwa, An interesting subseries of the harmonic series, Amer. Math. Monthly, 82 (1975) 931—933.

25. ——, Some convergent subseries of the harmonic series, Amer. Math. Monthly, 85 (1978) 661—663.

26. D.I.A.Cohen, An explanation of the first digit phenomenon, J.Combinatorial Theory (A), 20 (1976) 367—370.

27. R.A.Honsberger, Mathematical Gems II, Math. Assoc. of America, Wa

shington, D.C, 1976.

28. L. Gillman (oral communication).

29. B. D. Craven, On digital distribution in some integer sequences, J. Australian Math. Soc, 5 (1965) 325—330.

30. P. Diaconis, The distribution of leading digits and uniform distribution mod 1, Ann. Prob, 5 (1977) 72—81.

31. R. Baillie, Sums of reciprocals of integers missing a given digit, Amer. Math. Monthly, 86 (1979) 372—374.

四、付比尼原理与狄里克雷原理

S.K.斯坦因

有些数学工具是有名称的，比如二项公式，正弦定律，毕达哥拉斯定理，但是有些数学工具数学家天天都在用，却连个名儿也没有，人们动用这些工具时不象是在用锄头凿子，而更象是受鬼使神差一样。

我想在下面几页中使用一种工具，讲一讲它的几个应用，让它脱离不应有的无名状态，给它命个名：付比尼原理。它经常是和狄里克雷原理一起用，所以我先讲狄里克雷原理。

狄里克雷原理(专业领域里叫"抽屉原理")最简单的表达出来是：**如果你的信封比抽屉多，你把每个信封放进抽屉中，那么到头来总有一个抽屉里装了不止一个信封**.这很接近于下面这个更一般的形式：

如果每个抽屉平均装a个信封，那么有一个抽屉至少装a个信封。同样，一定有一个抽屉至多装a个信封。

现在举一个狄里克雷原理的简单应用，设想有一个容纳2000人的礼堂。运用狄里克雷原理我们知

道礼堂坐的人中至少有6个人同一天过生日。因为就我们这个情形来说等于是有366个抽屉(包括二月29日)和2000个信封。每个抽屉平均装信封2000/366≈5.5个。因此一定有至少5.5个人——因而也就是至少6个人过某个相同的生日。同理，一定至多有5.5人——因而也就是至多有5人过某一个相同的生日。

还有一个礼堂问题以后也会有用，试想象一个矩形礼堂，它有七行九列(图9)：

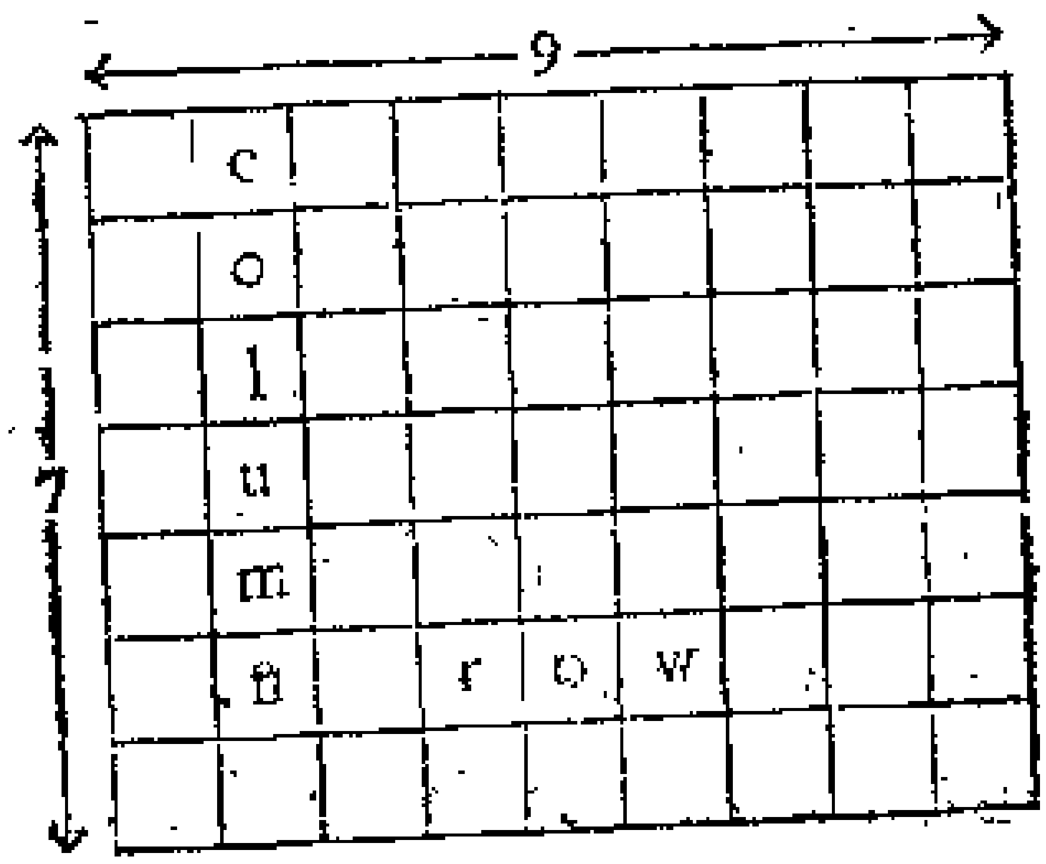


图9

行与书的地脚平行；列与书的侧面平行。(注意这两个术语，下面要用到。)

设想礼堂里坐了30个人。每行平均30/7≈4.3人，按狄里克雷原理，某行至少有五人；同理，某行至多坐了四人。考查每列的情况，对于列也会得出类似的结论。但是我们现在姑且把注意力转向付

比尼原理吧，这个原理在这以前连个名儿也没有哩。

现在设想你要数一数这个7×9的小礼堂中坐了多少人，如图10所示涂阴影的方格表示坐了人的。

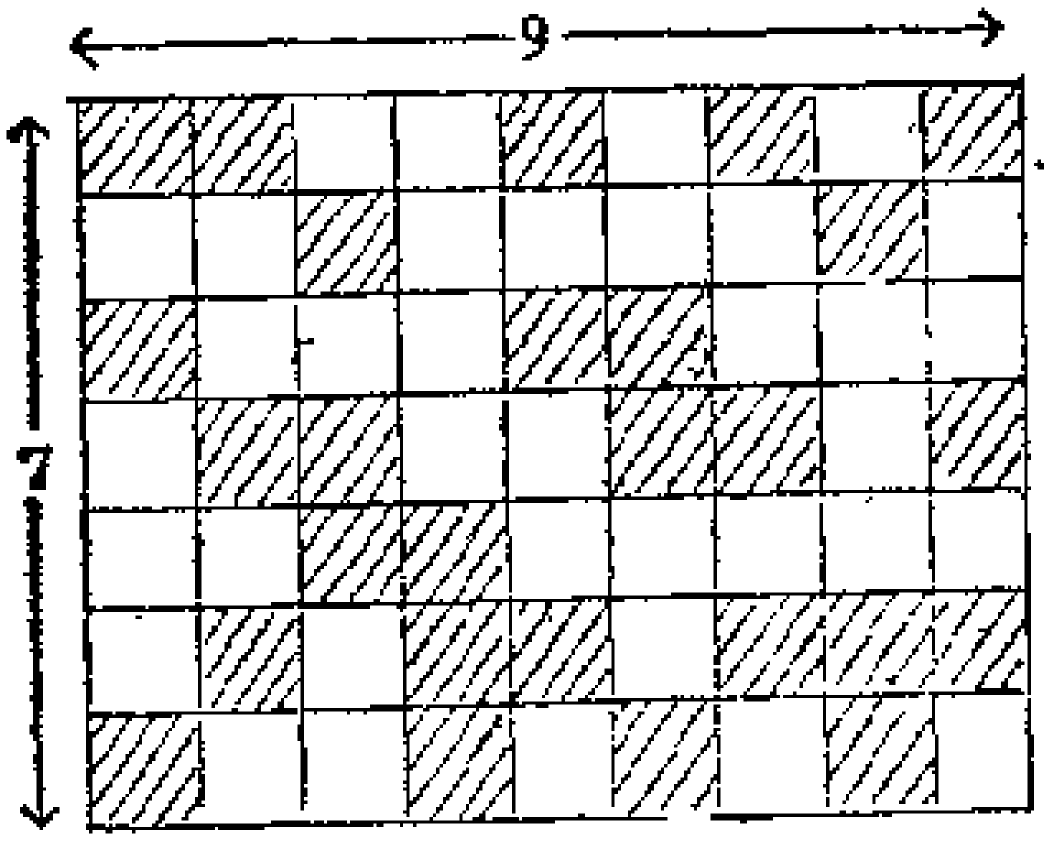


图10

一种"数房子"的方法是数每行的数目，然后把结果加起来，(从最上面一行数下来)得

$$5+2+3+5+2+6+4=27.$$

但是另外还有个方法：先数每列里的个数，瞟一眼就看到了每列都是坐的三个人，因为共有九列，所以坐了

$$9\times3=27$$

人，两个方法对比就把付比尼原理说清楚了：

你用两种方法去数一个集合，数出的结果是一样的，具体到礼堂问题，你数每列与你数每行得出的有人坐的位子都是一样多。

付比尼是因“付比尼定理”而出名的，这个定理说出了计数(微积分学的“积分”)的顺序不影响结果的相当一般的条件。

我们以后还会用到付比尼原理的一般形式，这个形式若用礼堂坐人的问题来叙述则可说成：

行数×每行坐人数=列数×每列坐人数

总之，方程两边都是说的有人坐的坐位的总数。

大家在中小学就知其然不知其所以然地学过2×3=3×2，那时就在用付比尼原理了，因为2×3的意思就是“数一数两个三元集中有多少个x”(图11)：

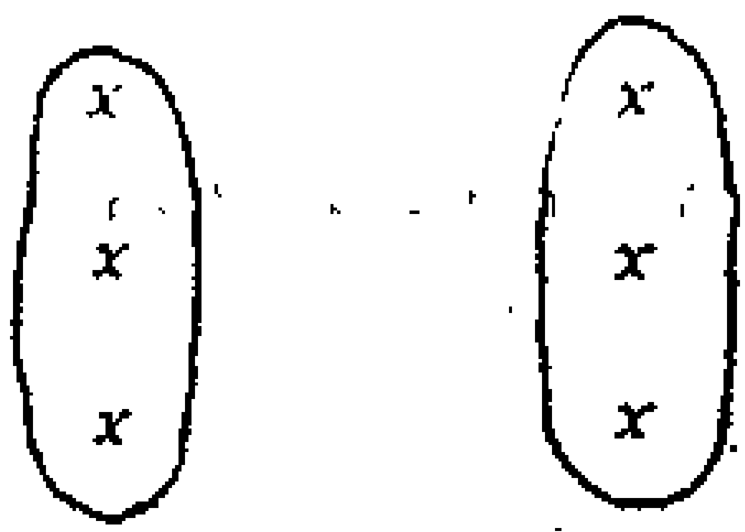


图11

不按列而按行来数，那就是3×2个x了(图12)：

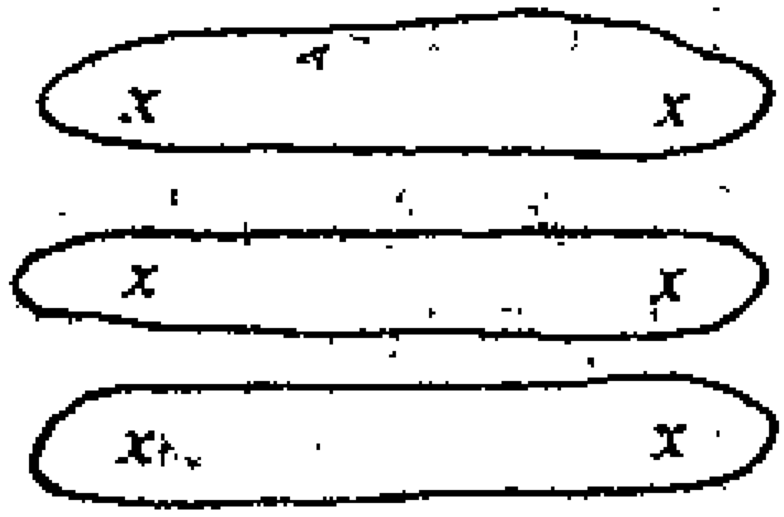


图12

所以$2\times3=3\times2$，因此，正整数的乘法交换律不过是付比尼原理的一个特例罢了。

1.数论中的付比尼原理 作为解释付比尼原理的更重要的例，我们再选一个关于数的因子问题来看一看，对于每个正整数k，令$d(k)$表示能整除k的正整数的个数。例如$d(1)=1$，$d(2)=2$，$d(3)=2$，$d(4)=3$，$d(5)=2$，$d(6)=4$；对任意素数p，$d(p)=2$，$d(k)$值的摆动幅度可以很大。例如，

$$d(41)=2,\ d(42)=8,\ d(43)=2.$$

即便如此，人们还是可以问："n增大时$d(k)$的前n个值的平均值变化情况如何？"更准确地说，当n增大时

$$\frac{d(1)+d(2)+\cdots+d(n)}{n}$$

如何变化法？

我们用付比尼原理从另一个角度来观察和数

$$d(1)+d(2)+\cdots+d(n),$$

并证明它近似地等于

$$n\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right),$$

首先以$d(6)$为例，它就可以看成j能整除6的数对$(j, 6)$的个数、所以$d(1)+d(2)+\cdots+d(n)$是正整对(j, k)的对数，这里j整除k，k小于或等于n，为了更直观地看这个数对的集合，可以想象有

一个n行n列的礼堂。令$(j,\ k)$中的j对应于一个行，k对应于一个列，如图13所示。若j整除k，我我们就把j与k交会之处当作"坐了人"，用涂阴影来表示。所以第2行与第6列交会处的位子"坐了人"，第3行与第7列交会处的坐位没有坐人，如图14所示。注意，$d(k)$是第k列"坐了人"的坐位的个数。

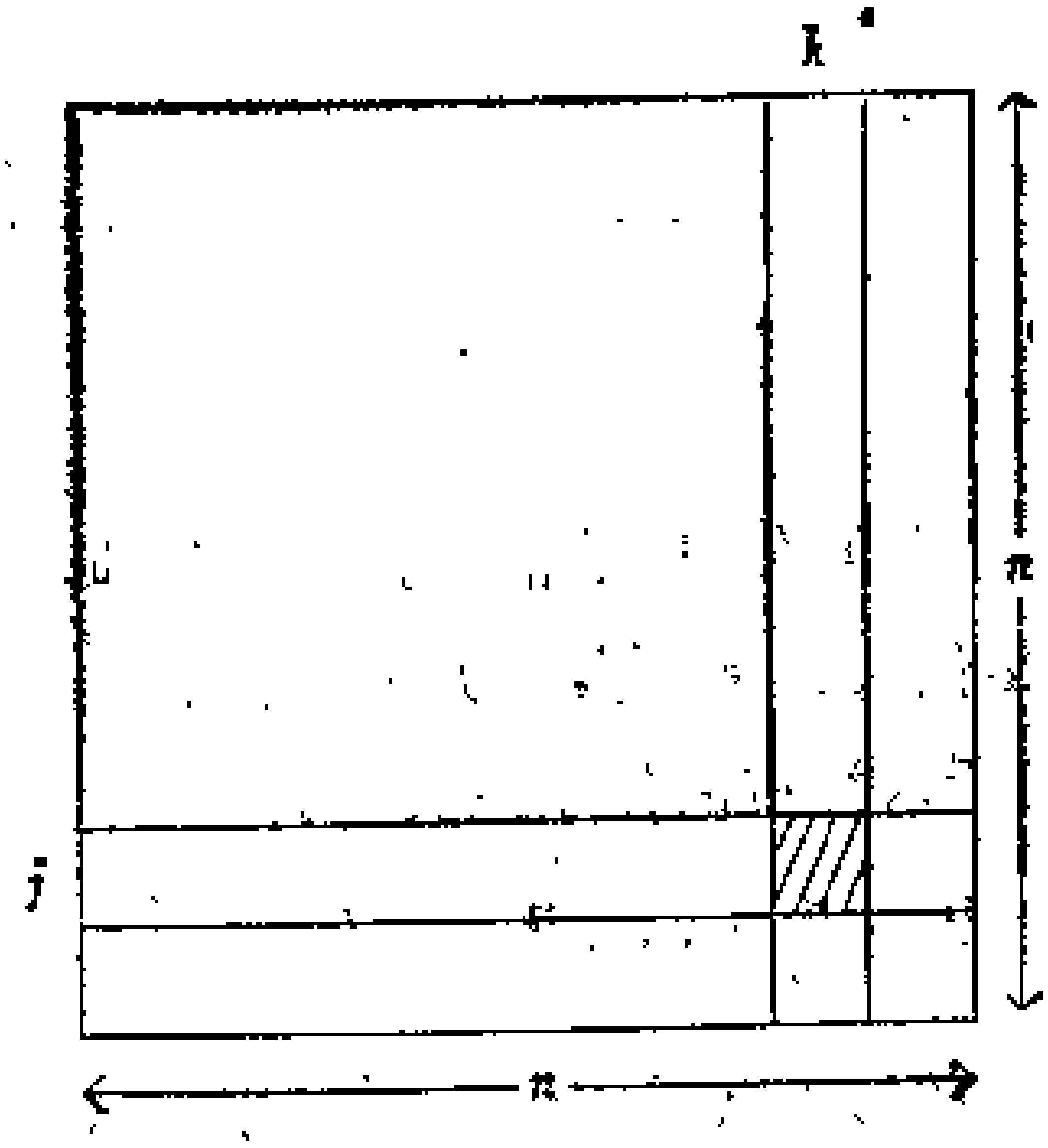


图13

和数$d(1)+d(2)+\cdots+d(n)$是一列一列地数出的"坐了人"的坐位的总数，利用付比尼原理，我们来一行一行地数"坐了人"的坐位，作为例子，第三行有多少个坐位坐了人？换句话说，有多少正整数k，$k\leq n$，能被3整除？这就是要求1与n之间的3的倍数的个数。这个数目大约是$n/3$，因为

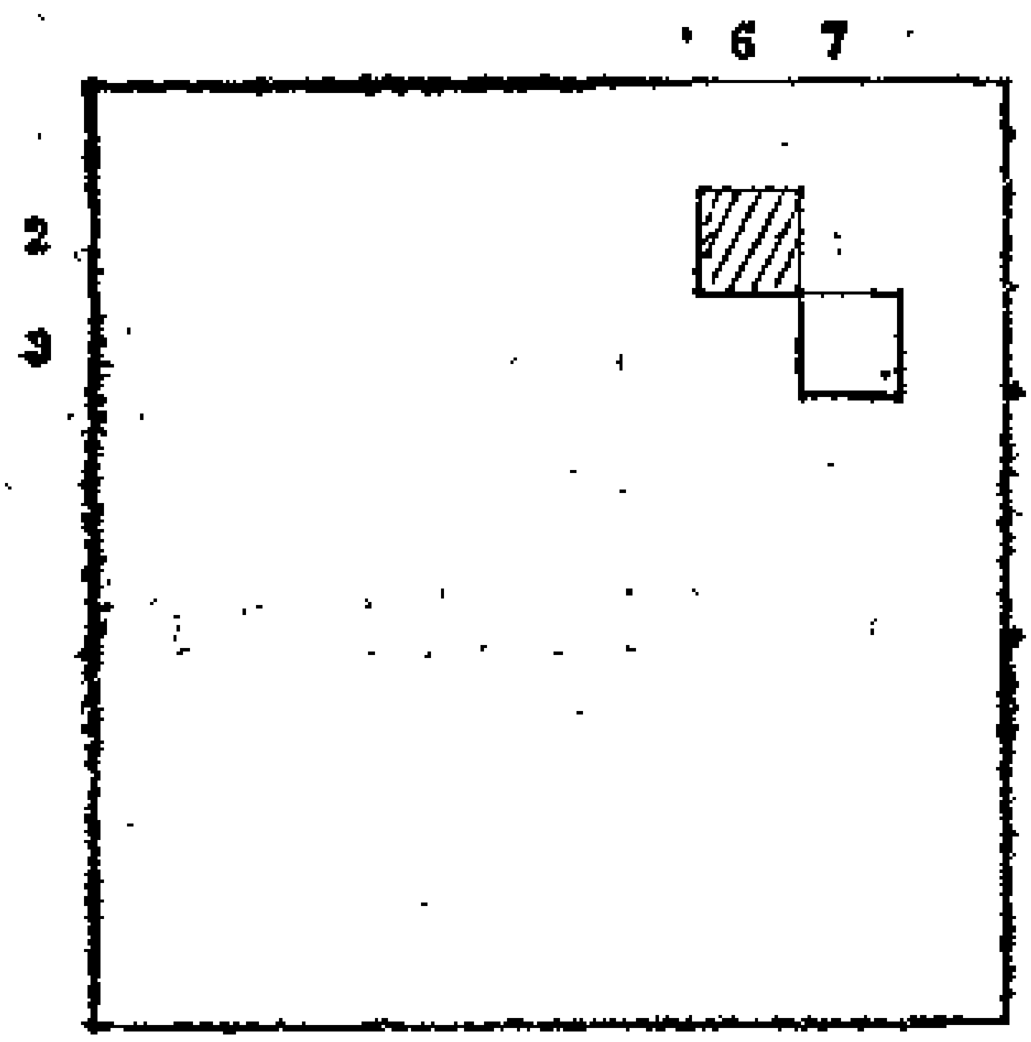


图14

每逢第3个整数便是3的倍数。更一般地，第 j行被人占的坐位大约是n/j。(实际上是在n/j与$(n/j)-1$之间。)所以被占的位子总共约为

$$\frac{n}{1}+\frac{n}{2}+\cdots+\frac{n}{n}=n\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right).$$

比起和数$d(1)+d(2)+\cdots+d(n)$来，这个和数是容易估计得多了。微积分学中已计算过，和数

$$\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}$$

近似地等于以e为底n的对数，$e\approx2.718$。(关于这个对数的讨论，参看练习3，所有的电子计算器都有$\log_e x$键。)所以$(d(1)+d(2)+\cdots+d(n))/n$大约等于

$$\frac{n\log_e n}{n} \doteq \log_e n.$$

这个式子表明，尽管$d(k)$摆动幅度很大，但是随着k增大，k的因子数目平均看来终究是增大的。

为以后引用方便起见，我们把这个结论用定理的形式记录下来，术语“大约”（近似地”）在后面的练习1、2、3中有解释。

定理1. 前n个正整数的因子的平均个数

$$\frac{d(1)+d(2)+\cdots+d(n)}{n}$$

近似地等于

$$\log_e n.$$

付比尼原理的第二个应用是在拓扑学中。

2. 拓扑学中的付比尼原理 考虑完全覆盖球面的一张地图，更准确地说是想象球面被分割成一些曲线所围成的区域，地图的一个小部分的样子如图15所示。

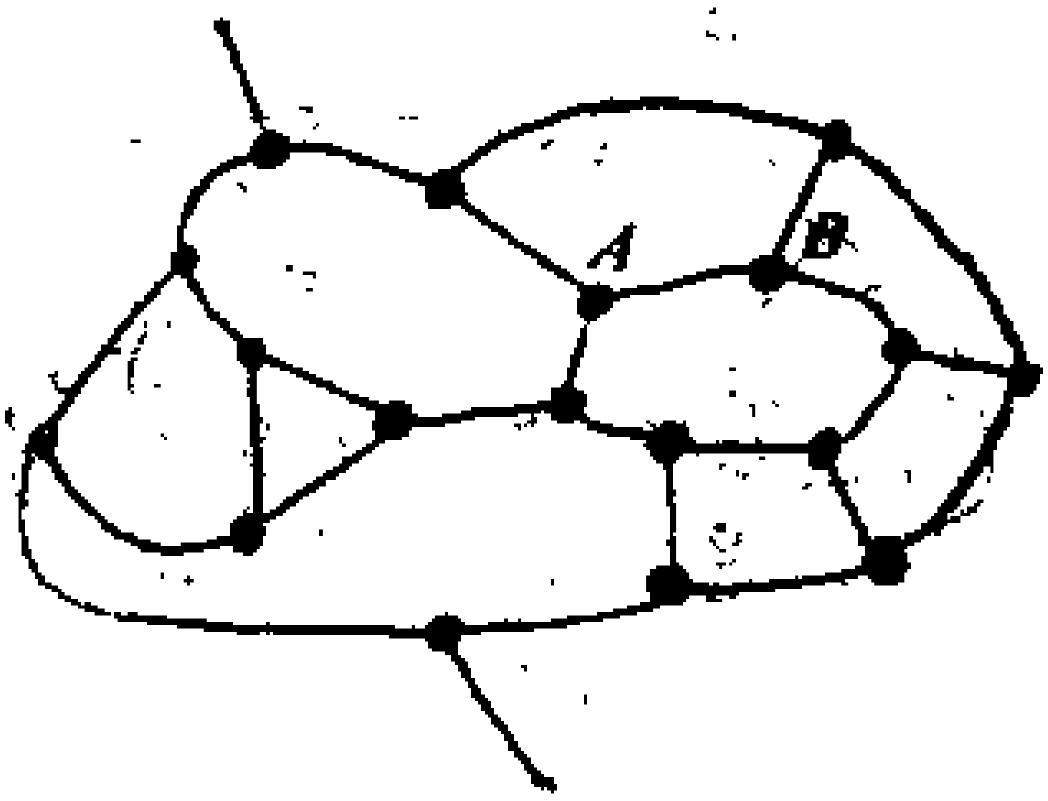


图15

假定在一点相会的至多只有三个区域。称三个区域的确交会的那个点为顶点。称连结两顶点而为两个区域的公共边界为棱。图15中粗黑点都是顶点，AB是一条棱。

就地图来说，令V为顶点数，E为棱数，R为区域数，十八世纪的欧拉证明了联系这三个数的方程

$$V-E+R=2.$$

利用这个关系式以及付比尼原理和狄里克雷原理我们来建立下面的定理，这里的表达方式不太正规。

定理2.在所说的地图上的一切区域中至少总有一个区域的边界至多有五条棱。换句话说，至少有一个区域是三角形，或四边形，或五边形。

为证明这个定理我们来证明每个区域的棱的平均数小于6。

我们令a表示区域的棱的平均数，考虑地图上

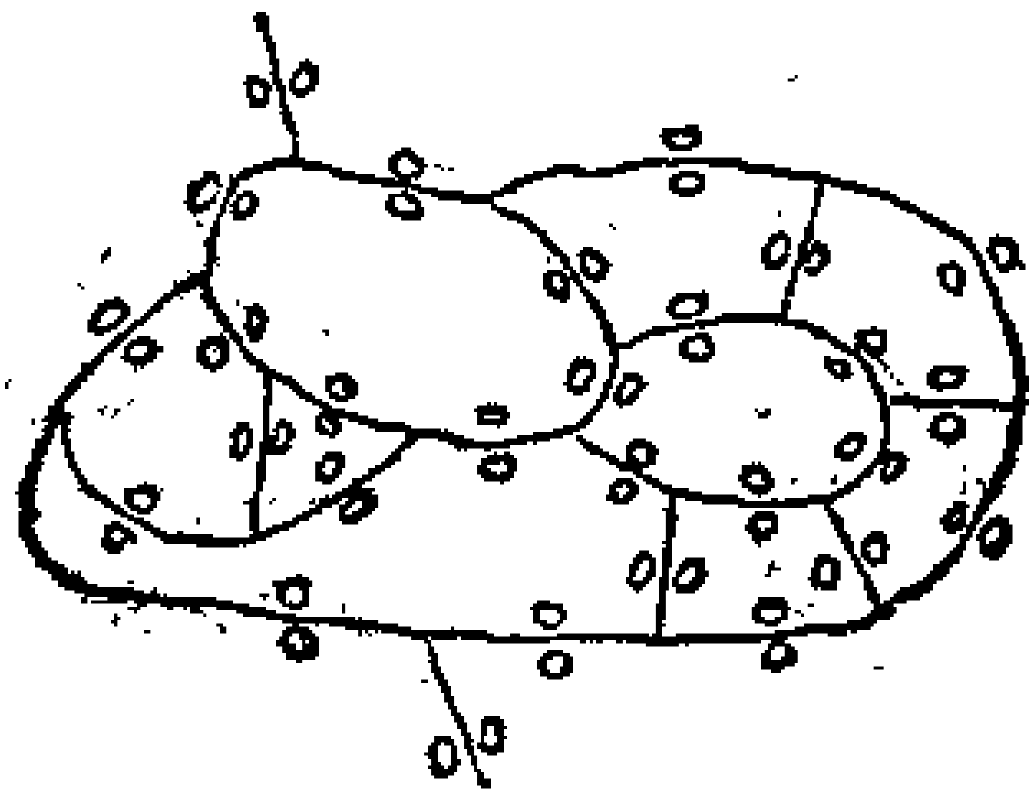


图16

的一个区域，在区域内靠近每条棱放一枚卵石．对每个区域都这样放卵石，于是卵石的分布就如图16所示，每条棱旁有两枚卵石．现在用两种方法来数卵石，首先卵石个数等于乘积

$$Ra,$$

因为有R个区域，每个区域中卵石平均数是a但，是另一方两在E条棱的每一条的两侧各有一个卵石，所以卵石总数为$2E$。于是

$$Ra=2E. \qquad (1)$$

接着我们来求联系V与E的方程．为此把注意力放到每个顶点上来，在交于顶点的三条棱的每条棱上靠近顶点的地方放一枚卵石，完成这项任务后就得到如图17的卵石分布情况。

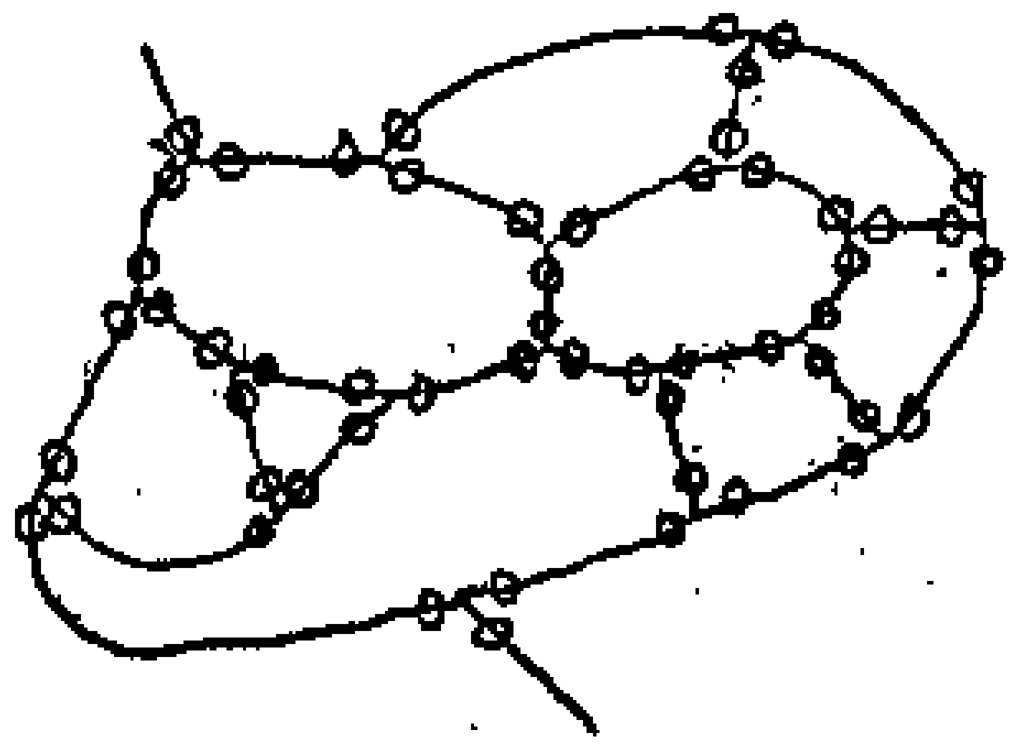


图17

注意，每条棱上有两枚卵石，这样一来卵石总数就

是$2E$，由于卵石总数又等于$3V$，故

$$3V=2E. \tag{2}$$

由(1)、(2)得

$$R=\frac{2}{a}E \qquad \text{和} V=\frac{2}{3}E$$

故欧拉方程$V-E+R=2$变为

$$\frac{2}{3}E-E+\frac{2}{a}E=2,$$

亦即

$$\left(\frac{2}{3}-1+\frac{2}{a}\right)E=2.$$

所以必定有

$$\frac{2}{3}-1+\frac{2}{a}>0,$$

因而

$$\frac{2}{a}>\frac{1}{3}$$

即 $$a<6 \tag{3}$$

按不等式(3)，每个区域的棱的平均数小于6. 所以根据狄里克雷原理一定有一个区域至多只有五条棱，应用这个定理可以证明球面或平面上的任意地图都可以用五种或不到五种颜色来着色，使得具有公共边的任意两个区域颜色不相同，请参看[5]，353—355页.

顺便说一句，前面的证明中既用了付比尼原理

也用了狄里克雷原理，这两条原理就象榔头与凿子一样总是形影不离，相得益彰，下面的例子更生动地说明了这一点：付比尼原理首先确定平均值大（或小）；狄里克雷原理随后据此作出关于此平均值的结论。

我们的第三个例题选自组合论，这是主要研究有限集的一门学问，这个例子在不同情况下利用了付比尼与狄里克雷两条原理。

3．组合论中的付比尼原理 我们从n^2个格的$n\times n$方阵开始，方格暂且空着。若$n=3$，我们就从图18开始：

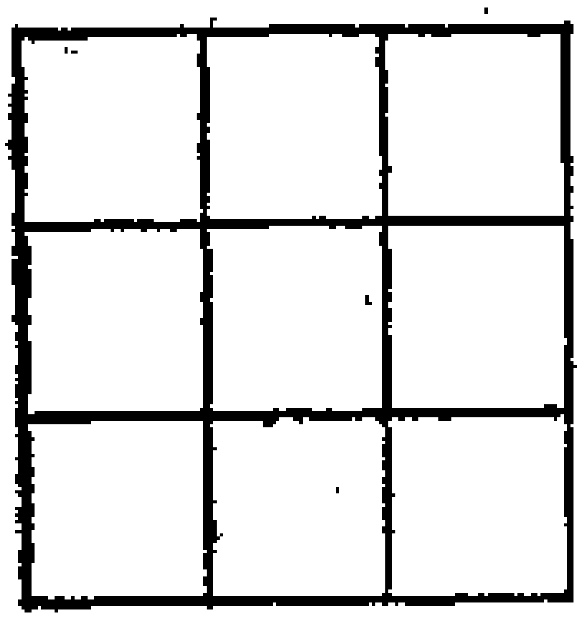


图18

在这种$n\times n$方阵中的n个格子的集合叫做横截的，如果其中任何两个格子都不在同一行也不在同一列，很容易检证，$n=3$时有六个横截集合，如图19所示。

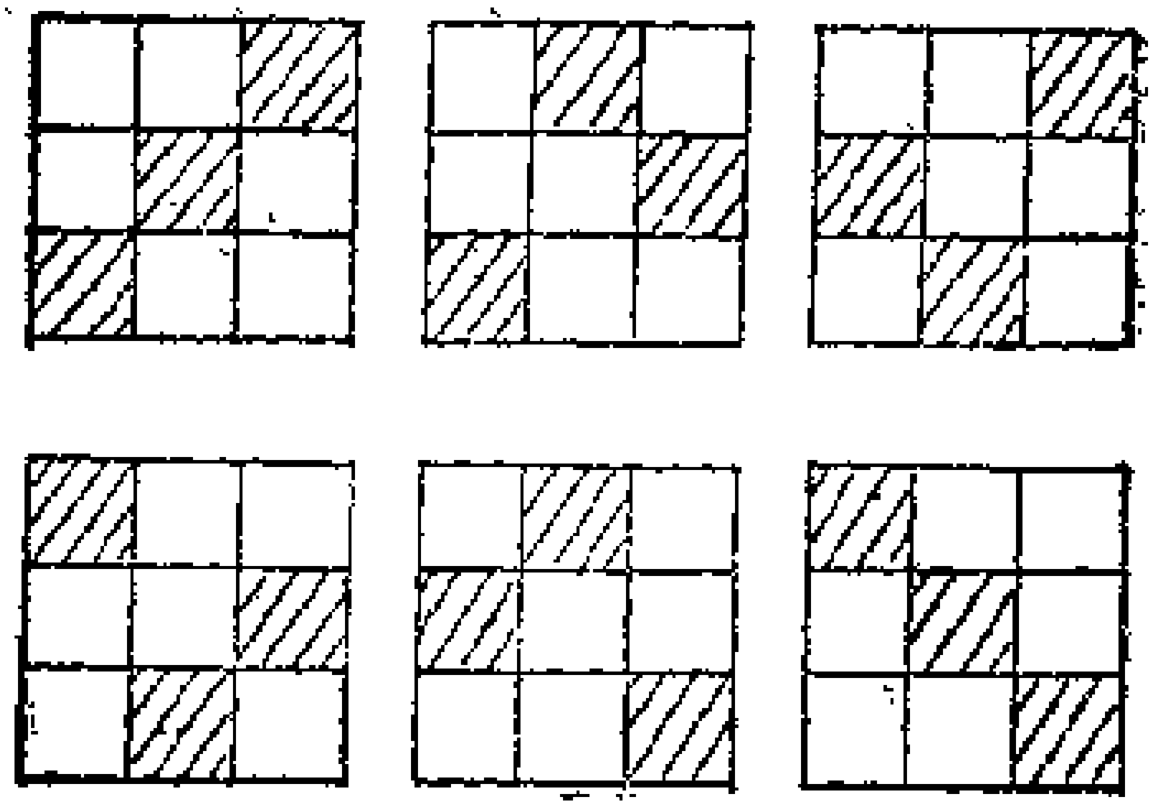


图19

$n\times n$方阵中有多少横截集合？（我们需要知道它）为计算此数我们首先注意第一行的那一个格子可以有n个选法，它一选定（“涂上阴影”）则第二行上的那一个格子便只有$n-1$种选择余地了，因为第二个格子必须在与选定的第一个格子不同的列上．所以横截集合中的属第一、二两列的那两个格子有$n\cdot(n-1)$种不同的选法．同理，从第三行上再选一个格子便只有$n-2$个选法，如此继续最后得到$n\times n$方阵中有

$$n(n-1)(n-2)\cdots(3)(2)(1)$$

个横截集合．用此公式有$3\cdot2\cdot1=6$，完全符合上面提到过的六个横截集合，前n个正整数的乘积，即$n(n-1)(n-2)\cdots(3)(2)(1)$这个数叫“$n$阶乘”，表示成$n!$．例如，$4!=4\cdot3\cdot2\cdot1=4\cdot(3!)=4\cdot6=24$．同样$5!=5\cdot(4!)=5\cdot24$

=120.

现在想象n^2个格子的每一个都包含一个数，另外再假定出现在一个格子的每个数只在两个格子中出现．于是n^2可被2整除，所以n是偶数。图 20是$n=4$的情形：

1	7	6	3
1	2	4	6
5	8	3	5
7	8	4	2

图20

请读者自己动手把 $n=2$，6，8的情形也画一画。

4×4的方阵中有几个无重复数字的横截集合。图21画出一个这样的横截集合：

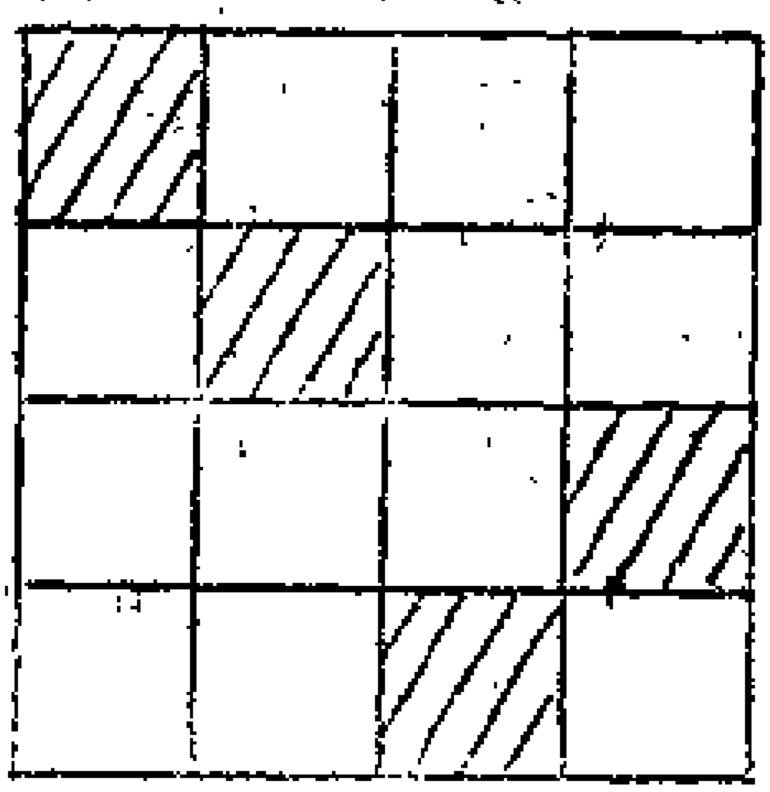


图21

你们还可以举出更多的例子，并验证6×6和8×8的情况，是不是一定有无重复数的横截集合呢？

你作试验时会注意到随机挑选出的大多数横截集合没有重复数，的确每个横截集合中不同数字的平均个数看来接近于n，当然每个横截集合至少有$n/2$个不同数了。（为什么？）

事实上$n=2$时不一定有无重复数的横截集合，例如下面这个2×2方阵的两个横截集合的每一个都有重复数字(图22)：

1	2
2	1

图22

下面的定理表明2×2这种情况是一般规律的例外。

定理3. $n\geq 4$**为偶数，在每个数只出现两次的任意**$n\times n$**方阵中有一个无重复数的横截集合。**

注意，如果填有同一数的两个格子在同一行或同一列上，那么每个横截集合就无重复数，问题出在填有相同数字的两个格子不在同一行或同一列上，因此如图23所示的包含两个格子的任意一个横截集合天然就有了重复数。

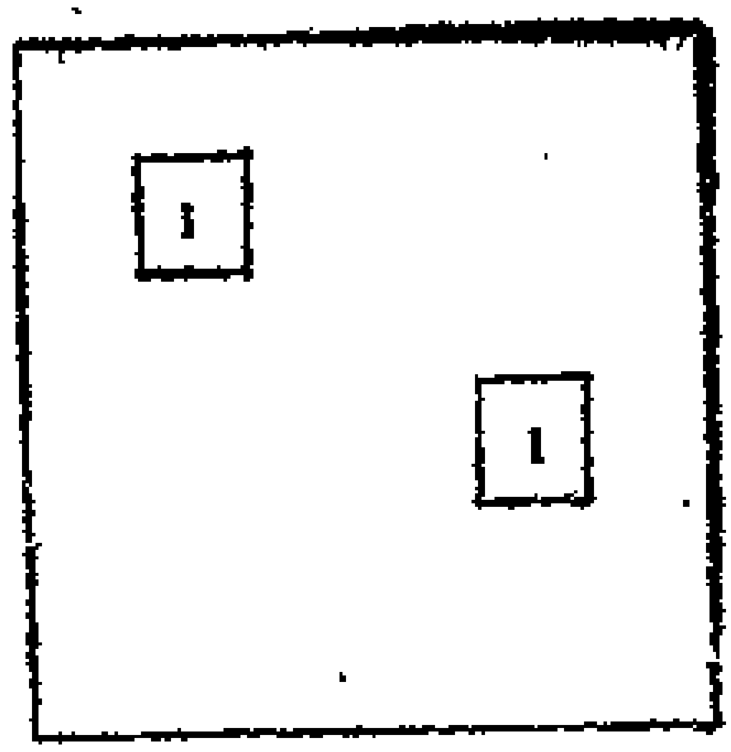


图23

我们把选自不同行和不同列且包含同一个数的一对格子称为奇异对．不含奇异对的横截集合没有重复数．因此我们要证的是，有一个横截集合不含奇异对．为此目的我们来考查在$n!$个横截集合中奇异对的平均数目．

假定我们的$n\times n$方阵有p对奇异对．（于是p就在0到$n^2/2$之间，视所给方阵而定．）画一个$n!$行p列的礼堂，让每行对应一个横截集合t；每列对应一个奇异对s，如图24所示

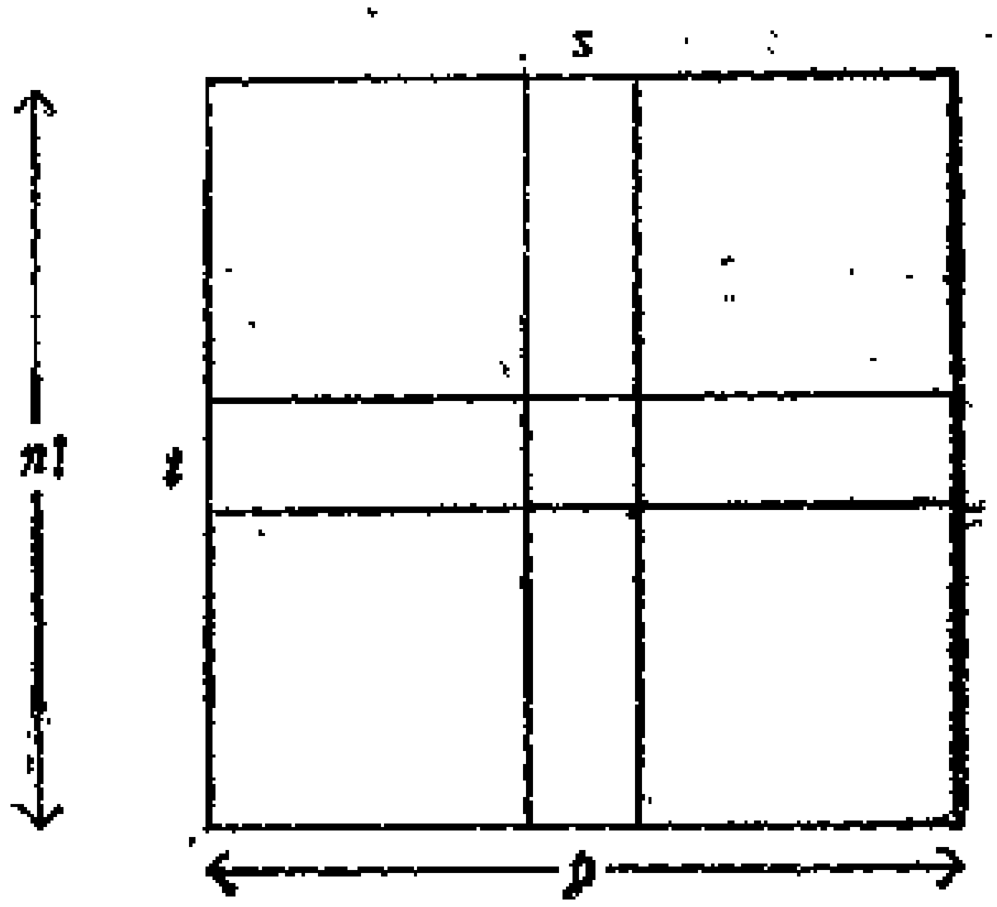


图24

位于第 t 行第s列的坐位当横截集合 t 含奇异对s时就是“被人占”了的坐位.

现在用不同方法来数被占坐位总数，首先，这个总数是

$$(n!)(\text{每个横截集合中奇异对的平均数}). \tag{4}$$

但按付比尼原理它又该等于

$$(p)(\text{含一已知奇异对的横截集合的平均数}) \tag{5}$$

幸好后面的平均数容易计算，无论你挑那一对奇异对，正好有$(n-2)!$个横截集合含这一对。(为什么？)所以合并(4)与(5)我们得到方程

$$(n!)(\text{每个横截集合中奇异对的平均数}) = (p)(n-2)!\ .$$

记得曾有$p\leq n^2/2$，所以上式两端除以2得

每个横截集合中奇异对的平均对数

$$\leq\frac{n^2}{2}\cdot\frac{(n-2)!}{n!}.$$

但是

$$\frac{(n-2)!}{n!}=\frac{(n-2)!}{(n)(n-1)(n-2)!}=\frac{1}{(n)(n-1)}.$$

结果每个横截集合中平均奇异对对数小于或等于

$$\frac{n^2}{2}\cdot\frac{1}{(n)(n-1)}=\frac{n}{2(n-1)}.$$

最后，由于$n \geqslant 4$，所以商$n/[2(n-1)]$至多等于2/3，因而小于1.

由于每个横截集中含的奇异对平均数小于1，根据狄里克雷原理必定有一个横截集合所含奇异对少于1.简单地说就是有一个无重复数的横截集合，定理证毕。

你读完证明后会发现定理的依据就是希望一个随机选取的横截集合中重复数很少，随机选出的东西可能相当有效这一想法在下面这个从组合论中选出的例子里得到了充分的发挥。

4. 有效的随机性 雨点开始落到人行道上时总是很快地落满人行道而很少重叠，大自然的这一貌似无序的过程事实上很有效，而人类一般却是被迫有序地完成自己的任务，毕尽信笔涂鸦鬼画桃符还是很难的。同样，要造一个完全随意的数字符号序列必须用到很深奥的数学，下面的例子表明有些时候乱猜是很不错的猜测，大自然洒水在大地上是很可靠的。

1957年C.A.罗杰斯应用“随机性”确证了空间薄覆盖的存在，这个覆盖用了已知集的平移。即使在球平移的情形罗杰斯的薄覆盖也是高维空间的最佳覆盖，因为所有更明显的已知方法作出的覆盖都是重叠太多。

需要集合论的几个概念，尽管是很常见的，还

是停下来复习一下好些。q个集合S_1，$S_2\cdots$，S_q的并集是由至少q个集中某一个集的那些元素组成的，表示成

$$S_1\cup S_2\cup\cdots\cup S_q$$

比如以q为5且每个集S_i有3个元素为例，那么它们的并集至多有15个元素，但至少有3个元素。

集合S_1，S_2，$\cdots$，S_q的交集是全部q个集合的公共元素的集合，表示为

$$S_1\cap S_2\cap\cdots\cap S_q,$$

交集有时还可能是空集，即不含元素的集合。

令集合S位于集合X之内，那末属于X但不属于S的元素组成的集合叫做S在X中的补集，写成

$$X-S.$$

图25是上面三条定义的图解：

阴影部分是$S_1\cup S_2\cup S_3$

阴影部分是$S_1\cap S_2\cap S_3$

阴影部分是$X-S$

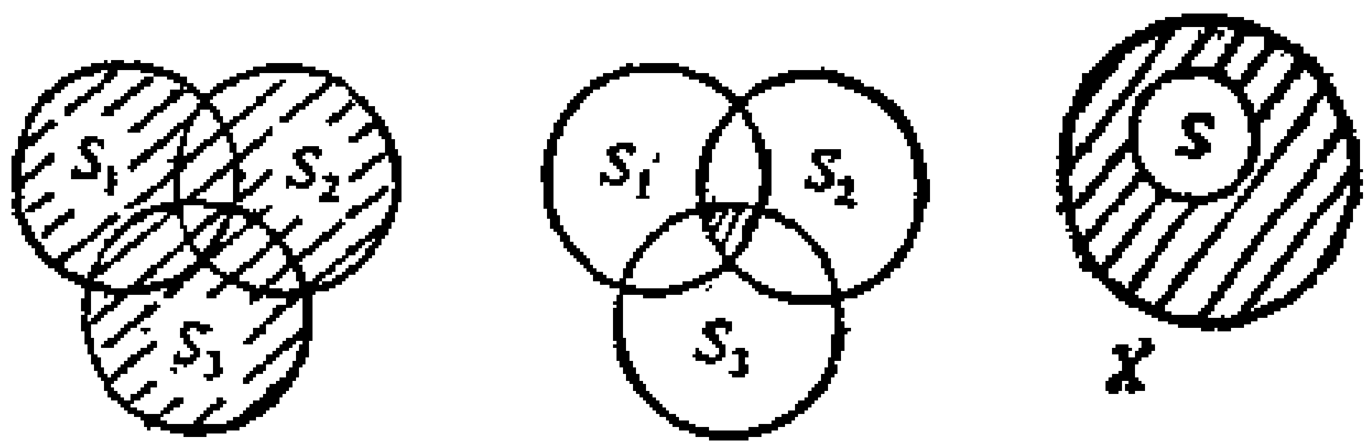


图25

读者可以验证并、交、补的下列关系：

$X-(S_1\cup S_2\cup\cdots\cup S_q)$

$$=(X-S_1)\cap(X-S_2)\cap\cdots\cdots\cap(X-S_q).$$

每个S_i都是假定在X内，把X看成是人行道，把S_i看成是X的被雨点S_i打湿了的部分．这对以后很有用．$S_1\cup S_2\cup\cdots\cup S_q$是$q$个雨点打湿了的地面，所以上面等式的意思就是“人行道的干路面就是每滴雨未打湿的部分之交”．

现在回到例子本身来，又来考虑分成n^2个小格的$n\times n$方阵，每个小格内填入1，2，…，n中的一个整数，而且任何一行或任何一列中不准有整数重复．这样排列出来的阵列叫n阶**拉丁方**，用符号

$$r\circ s$$

表示第r行与第s列相交处的小格中的元素，图26上是一个6阶拉丁方；注意，比如有$2\circ3=1$。

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	1	5	6	4
3	3	1	2	6	4	5
4	4	6	5	2	1	3
5	5	4	6	3	2	1
6	6	5	4	1	3	2

图26

我们来证明下面关于任意拉丁方的定理，

定理4　考虑一个n阶拉丁方和能整除n的一个正整数m．令$n=qm$．那么一定存在q行，使得这q

行的前m列中的整数集合的并集至少含有

$$n[1-(1-1/q)^q]$$

个不同元素

我们暂且停下来看一看，关于图26的拉丁方定理4告诉我们些什么。这时$n=6$，令$m=2$，它能整除6．于是$q=3$．所以定理4说的就是存在着3行，使得这3行的前2列至少包含

$$6\left[1-\left(1-\frac{1}{3}\right)^3\right]$$

个不同的元素。但是

$$6\left[1-\left(1-\frac{1}{3}\right)^3\right]=6\left(1-\frac{8}{27}\right)$$

$$=\frac{38}{9}=4\frac{2}{9}.$$

所以定理4确保我们能找到某三行使其头二列至少有5个相异元素，一看图26立刻就找到三行，它的头2列一共有5个相异元素，例如第1、2、4行便是．但是在图26中却没法找出3行，其头2列有6个互异元素，这一点顺便看一眼即可证实。

定理4一般地说了些什么呢？首先，$(1-1/q)^q$这个数小于1，我们列出它的几个数值来，准确到小数点后两位：

q	2	3	4	5
$(1-1/q)^q$	0.25	0.30	0.32	0.33

微积分学中证明了，随着q增加，$(1-1/q)^q$也慢慢地增加并趋近于$1/e\approx0.37$，所以定理4蕴涵的意思是，我们能够找到q行，在其头m列中它们总共至

少包含$n(1-0.37)$个相异的数，换言之，占整数1，2，…n的63%的数。

当然有时我们可能会获得比63%更好的结果，例如我们若能找到q行使得它们在前m列中的那部分根本不重叠，那么它们一共就会有$q\times m=n$个整数，这就是1，2，…，n诸整数的100%了．总之，无论如何63%已经够可以的了，因为这表明能找出q行，使得它们在前m列中重叠得并不太厉害。好，我们来看证明。

令X为整数1，2，…，n的集合，A为整数1，2，…，m的集合。第x行中位于前m列的元素是

$$x\circ 1,\ x\circ 2,\ \cdots,\ x\circ m.$$

为方便起见把它们组成的集合简写为

$$xA.$$

注意xA有m个元素；因此集合 $X-xA$ 有$n-m$个元素。

我们希望找出q行x_1，x_2，…，x_q，使得

$$x_1A\cup x_2A\cup\cdots\cup x_qA \qquad (6)$$

至少含有

$$n\left[1-\left(1-\frac{1}{q}\right)^q\right]$$

个元素。

为此目的，我们在1，2，…，n这n个数中随意挑选出x_1，x_2，…，x_q，并找出对所有可能的x_1，

x_2, …, x_q的选法(6)中的元素的平均个数。(把1, 2, …, n 这个集合看成是人行道路面，那末集合 x_1A, x_2A, …, x_qA 便是 q 个雨点，而并集 $x_1A\cup x_2A\cup\cdots\cup x_qA$便是$q$个雨点打湿了的那一部分人行道路面。)

为计算方便起见，从(6)的补集

$$X-(x_1A\cup x_2A\cup\cdots\cup x_qA) \tag{7}$$

开始算，而这补集又等于

$$(X-x_1A)\cap(X-x_2A)\cap\cdots\cap(X-x_qA). \tag{8}$$

(补集(7)代表人行道干的部分，它是每滴雨所未打湿的部分的交集(8)。)

设想有一个礼堂，对应于x_1, x_2, …, x_q的

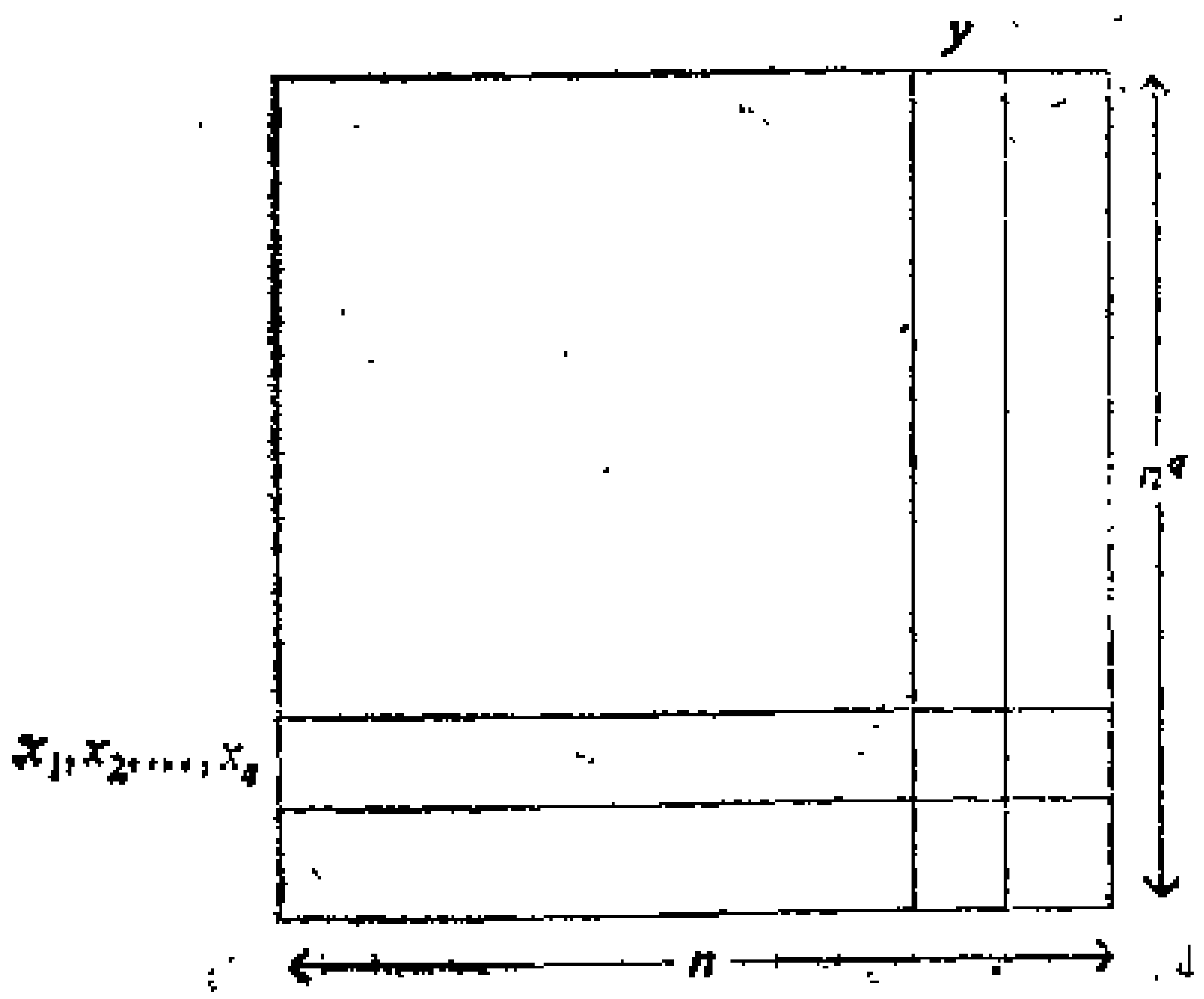


图27

每一种选法它都有一个行(见图27)，由于x_1有n个选法，x_2有n个选法，…x_q有n个选法，q次选择是独立的，所以总共有

$$n^q$$

行，礼堂有n列与1，2，…，n相对应，令y表示这n个数(即n列)的任一个，如图27所示。

如果

y在$(X-x_1A)\cap(X-x_2A)\cap\cdots\cap(X-x_qA)$中，我们称第$x_1$，$x_2$，…，$x_q$行与第$y$列交会处的坐位为"被占"坐位．记住，我们想得出的是被占坐位很少的一个行。

为此我们用两种方式来数被占坐位的总数，首先被占坐位总数为

行数×每行平均被占坐位数，

即

$$(n^q)(\text{每行平均被占坐位数}). \qquad (9)$$

但是根据付比尼原理，被占坐位总数也该是

列数×每列平均被占坐位数，

即

$$(n)(\text{每列平均被占坐位数}). \qquad (10)$$

我们考虑一个具体的列，y．行x_1，x_2，…，x_q与第y列交叉处的坐位什么时候是被占的？换句话说，什么时候y属于$(X-x_1A)\cap(X-x_2A)\cap\cdots\cap(X-x_qA)$？而这个问题又等于是，"什么时

候 y 不属于 x_1A，且不属于 x_2A，…，不属于 x_qA？”

我们具体地来考虑 x_1 有多少种选法能使 x_1A 不含 y。由于数 y 在 $n\times n$ 拉丁方的前 m 列中出现 m 次，所以 x_1 有 $n-m$ 个选法使得 x_1A 不含 y。因而对于每个 y，使得 y 不在 x_1A 中，不在 x_2A，…不在 x_qA 中，x_1，x_2，…，x_q 的选法个数为

$$(n-m)^q.$$

所以对每个列 y，有 $(n-m)^q$ 个被占坐位；(10)变成

$$n(n-m)^q. \tag{11}$$

合并(9)与(11)得

$$n^q(\text{每行平均被占坐位数})=n(n-m)^q,$$

所以

$$\begin{aligned}\text{每行平均被占坐位数}&=n\left(1-\frac{m}{n}\right)^q\\&=n\left(1-\frac{1}{q}\right)^q\end{aligned}$$

遵照狄里克雷原理，存在一个 x_1，x_2，…，x_q 的选法使得(7)充其量具有 $n(1-1/q)^q$ 个元素，对于这个选法 $x_1A\cup x_2A\cup\cdots\cup x_qA$ 再少也有

$$n-n\left(1-\frac{1}{q}\right)^q$$

个元素，证毕。

虽然这个证明比前面那些证明都复杂些，但是

你若仔细研究一番就会发现这不过就是雨点乱滴现象与付比尼、狄里克雷二原理的结合罢了。

在信息论，几何学等不同领域里，用类似的，不过更复杂一些的道理已经得到一些成果。1948年C.E.申农创立了关于信息及其在噪声信道中以适当代码进行传输的数学理论，把雨点乱滴法与付比尼、狄里克雷二原理结合起来，申农证明了随机构造出的代码几乎总是能使信息以接近信道容量的大小进行传输，不幸的是这个结论并不保证任何具体的码具有这个特性。事实上为了寻求产生有用的、有效的代码的方法，数学家已经奋斗了三十年了。(参见[1]，p.173.)

5. 通向未知的路 后两个例子虽然是专门选来说明狄里克雷原理和付比尼原理的，但是却传达了关于数学方面的更普遍得多的一条真理：回答一个问题所得到的报偿往往却是受到更多的新问题的挑战。数学的未知境界扩张之快，绝非探索者追赶所能及；在数学的空间中有许许多多谜的银河，这些谜就像望远镜里看到的最奇怪的星星一样地令人困惑莫解。而这个宇宙的边疆是如此的遥远，惟有我们的好奇心和想象力才能越过它。

这儿有一些问题，都是立刻从上面两个定理可以引申出来的。

定理3所说的阵列中每个记号只出现两次，现在可以问：

如果每个记号最多出现两次结论又如何呢？

如果每个记号只出现三次呢？最多出现三次呢？

如果每个记号只出现n次呢？

定理4谈的是拉丁方，考虑的是q行中前m列，$mq=n$。但是如果mq不等于n，定理4又将如何推广呢？

这些问题有些在练习中讨论，有些是至今尚未解决的。

读者还可以提出更多的问题来。

使数学家的思想永保青春，使数学家富于探险精神，使数学家永远想到人类肉眼和思想所未曾探索过的世界去遨游，办法就是提出问题，回答问题。而在装着有用工具的背包中一定有两件宝贝工具——狄里克雷原理和付比尼原理。

练习

练习1到练习3是关于定理1的。

1.(a)计算$d(k)$当k从1到15时的平均值。

(b)把(a)的平均值与$\log_e 15$比较，结果如何？

2.(a)我们曾看到，小于或等于n的i其倍数的

个数大约为n/i．试证这个数在n/i与$(n/i)-1$之间。

(b)试推导$d(k)$从k等于1到n时平均值是在

$$\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n} \quad 与 \quad \frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}-1$$

之间。

3．微积分学证明了曲线$y=1/x$下，x轴上，从$x=1$到$x=n$的面积是$\log_e n$(图28).

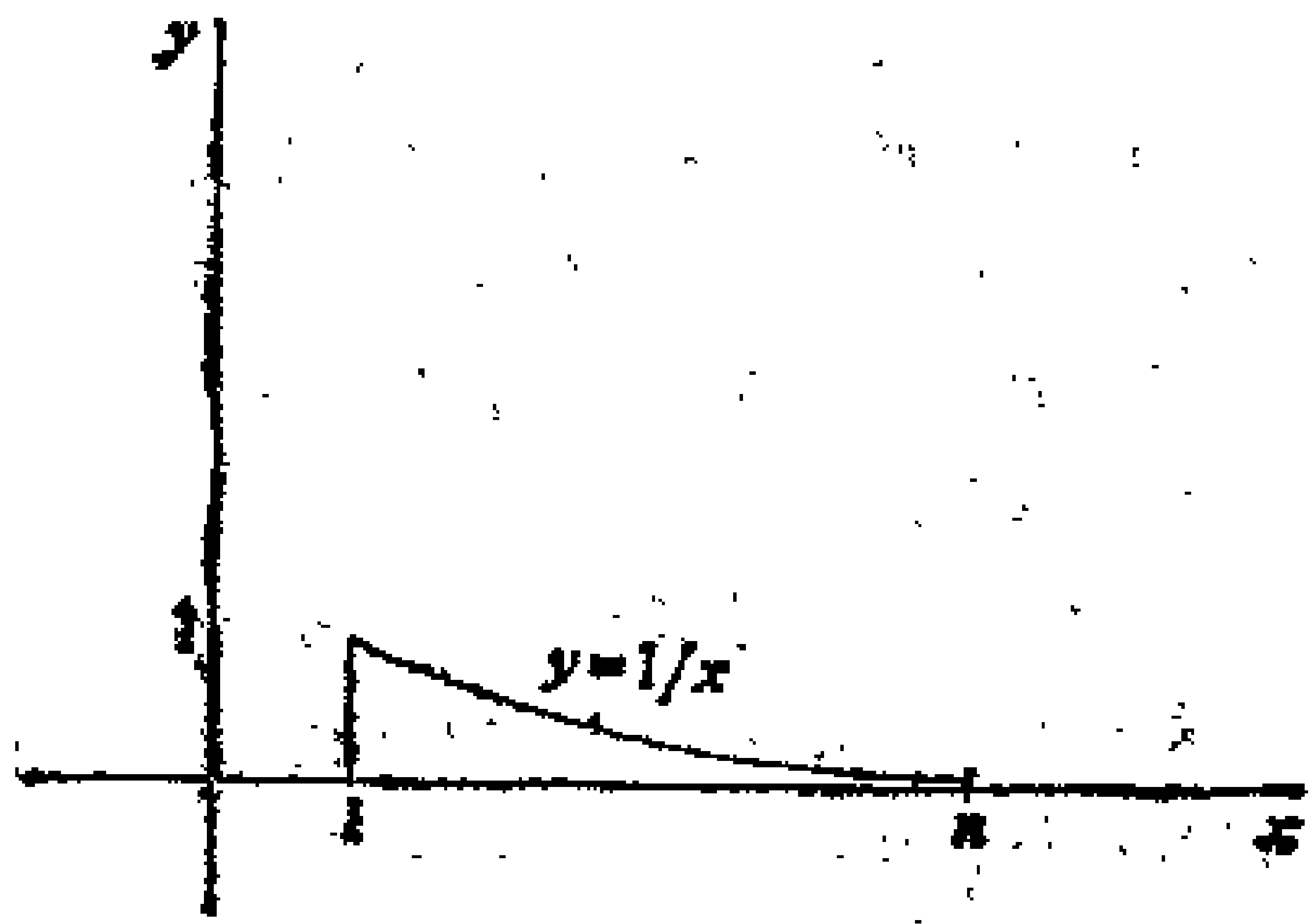


图28

(a)把和数

$$\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}$$

解释为某些适当的矩形面积之和，由此证明

$$\log_e n < \frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}.$$

(b)用同样办法证明

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}<1+\log_e n.$$

(c)推导：对于大n，比数

$$\frac{1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}}{\log_e n}$$

接近于1.

(d)推导：当 n 很大时，$[d(1)+d(2)+\cdots+d(n)]/n$与$\log_e n$之比趋近1.

练习4与5是关于球面上地图的。

4．定理2作的假定是在每个顶点处只有三个区域相会。试证明：如果换成假定在每个顶点至少有三个区域相会，那么仍然一定会存在一个三角形或四边形或五边形。

5．假定一个球的表面被分割成若干个3边形区域，于是每个区域的边界含有3条棱(边)，3个顶点。然后在每个顶点处写上A或B或C，试证明边界上三个字母全有的那些区域的数目是个偶数。提示：在每个三边形中放一个卵石在AB边(如果有AB边的话)。

练习6至10是关于定理3的题。

6．考虑每个符号至多出现两次的一个$n\times n$方阵。证明如果$n\geq3$，则存在一个横截集合，不含重复符号。

7．举出一个3×3方阵的例子来，这个方阵中

每个符号至多出现3次，然而每个横截集合都有重复符号。

8. 令$n\geq 4$. 考虑每个符号至多出现3次的一个$n\times n$方阵。试证存在一个横截集合，至少含有$n-1$个不同的符号。

9. 等n方阵是一个$n\times n$方阵，其中1，2，…，n的每一个都只出现n次。是否恒有一个横截集合，它至少有$n-1$个相异数？(参见[6].)

10. 考虑一个n阶拉丁方。每个横截集合的相异数平均有多少？(参见[6].)

练习11至19是关于定理4的。

11. (a)请解释为什么有

$$X-(S_1\cup S_2\cup S_3)=(X-S_1)\cap(X-S_2)\cap(X-S_3).$$

(b)请解释为什么有

$$X-(S_1\cap S_2\cap S_3)=(X-S_1)\cup(X-S_2)\cup(X-S_3).$$

12. "前m列"换成"后m列"定理4还成立么？

13. (见定理4.)考虑4阶拉丁方。试证有两行，使得这两行与前两列相交处的4个小格都不相同。

14. (见定理4.)令n为正整数，m与q为满足$1<m<n$与$1<q<n$的正整数。试证有q行能使它们与前m列相交处的mq个小格至少有$n(1-m/q)^q$个

相异数。

15．(见定理4.)令n为偶数，且$m=n/2$。令A为m个数的集合。试证存在两个元素x_1和x_2，使得

(a) $x_1A\cup x_2A$至少有$(3/4)n$个元素。

(b) $x_1A\cap x_2A$至多有$(1/4)n$个元素。

16．(a)在定理4的证明中，从整数1，2，…，n中排出了所有可能的x_1，x_2，…，x_q。会不会在某些情况下x_1，x_2，…，x_q相等呢？

(b) 证明，结论更强时定理4依然成立。这个更强的结论是：存在着q行，其前m列所含相异元素多于

$$n(1-(1-1/q)^q).$$

这是定理4的一个改进。

17．(见练习16b.)

(a) 证明8阶拉丁方中有两个行，它的前4列至少有7个相异整数。

(b) 有没有这样一个8阶拉丁方，它任何两行上的前4列都没有8个相异整数？

18．(a) 请对$n=8$，$m=2$，$q=4$叙述定理4．

(b) 还可以改进么？

19．(这个练习研究乱滴的雨点，是达到定理4的一个直观方法.)设想一块干路面，面积为一个平方单位。令q为正整数。考虑q个乱雨点的序列，

每个雨点打湿$\frac{1}{q}$平方单位的面积。(打湿部分面积一样但形状不必一样。为方便起见我们可以把雨点打湿的地方想成一个小圆。)再把雨理想化一点儿，不把打在边界上的雨滴算在内；这样一来每个雨点就是完全落在人行道路面上的了。

第一个雨点落下以后路面有$1/q$个平方单位的地面被打湿了，剩下$1-1/q$个平方单位的路面是干的。

(a) 任意一个第二雨点落下以后为什么你会认为平均有$(1-1/q)^2$平方单位路面是干的?

(b) 任意的第三个雨滴落下以后为什么你会认为平均有$(1-1/q)^3$平方单位路面是干的?

(c) 为什么你会预料到落下q滴雨以后平均说来$1-(1-1/q)^q$平方单位路面已经湿了?

(d) 为什么必定有q个雨点的一种分布，使得路面的$1-(1-1/q)^q$平方单位是湿的?

其余的练习是两个原理的进一步说明。

20. (给学过重积分的人做)。对积分

$$\int_0^1\left(\int_y^1 e^{x^2}dx\right)dy$$

进行估值。

21. (a) 给定27个整数，证明若用26除，则有两个数的余数是一样的。

(b) 给定27个整数，证明有两个的差是26的倍

数。

22.（这个题要用到练习21蕴涵的想法。）

(a) 证明：存在相异正整数m与n，使得41整除10^m-10^n。

(b) 证明：存在一个正整数e，使得41整除10^e-1。

(c) 为什么(b)中最小的e小于41？

顺便补一句，这个最小的e和1/41的十进小数展开式的循环节的长度是一样的。

23. 令n为正整数。

(a) 证明：存在n个都不大于$2n$的正整数，使得它们中的任何一个都不能整除另外任何一个数。（这个练习用不到讨论过的任何原理。）

(b) 证明：如果$n+1$个正整数都小于或等于$2n$，那么它们中存在两个数，其中一个能整除另一个。

24. 令$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n$是一个实系数$n$次多项式。设它有$n$个实根。

(a) 证明：存在一个根r使得$|r|\leq|a_1/n|$。

(b) 证明：存在一个根r使得$|r|\geq|a_1/n|$。

(c) 证明：存在一个根r使得$|r|\leq\sqrt[n]{|a_n|}$。

(d) 证明：存在一个根r使得$|r|\geq\sqrt[n]{a_n}$。

25. 考虑这样的n个城市，每两个城市之间有一条单行道相联。两条这样的路只相接于一个城

市。若要交叉，就修一个天桥。

(a) 证明：有一个城市，从它那里至少有$(n-1)/2$条单向道发出。

(b) 证明：有一个城市，向它那里至少有$(n-1)/2$条单向道汇聚。

26. 设想整个xy平面分成了许多个区域，每个区域以曲线或折线为边界。假定任意有限多个这种区域都可以至多涂上5种颜色，使得有公共边界的区域颜色不同。证明：全部这种区域都可以至多涂上5种颜色，使得有公共边界的区域颜色不同。

27. (给学过曲率半径的人做.)长度为L的一条曲线是一个凸区域的边界。(一个区域中如果任意两点连成的弦完全在该区域内，则称此区域为凸区域。)假设在这条曲线的每一点定义了曲率半径，且此半径沿曲线连续变化。证明：这曲线上有一点，其曲率半径为$L/2\pi$。

28. 设想有一个圆饼分成了n个全等的扇形块，n为偶数。令$m=n/2$. 在m块的每块中插入一只烛。证明：能够把饼旋转到一个新位置，使得与原始位置匹配的饼块数不超过$m/2$. 所谓"匹配"是指两块有一只烛或不是两块都有一只烛。关于这个结果在编码理论中的意义参见[9]，$p.348$. 提示：参看练习15。

29. 考虑一个$n\times n$方阵，其中1，2，…，n的每

一个符号出现n次。证明：有一行或一列至少含$\sqrt{n}$个相异符号。（〔7〕，pp.573－574中可以找到一个证明。

30．这里有一个几年前口头传播的问题。当时没有解；也许现在仍未解决。设想有一正方形木地板，它的边是南北走向和东西走向。假设有一些钉子从地板上突出来了，其中一颗在西南角。你有取之不尽的各种大小的矩形地毯。不消说你是想尽可能多地盖住地板，但却有些限制。首先，地毯边必须与地板边平行。其次，在地毯的西南角必须有某一颗钉子。最后，地毯不许与钉子或另一地毯重叠。问是否可能覆盖半个地板？

当你不在意地，随机地，合乎规定地铺地毯时，地毯平均地说来至少盖住了一半地板，这情况会出现么？换个说法，地板的任一给定位置能否被遮住一半？如果有这种可能性，付比尼原理和狄里克雷原理就可以把问题解决了。我要让读者看看上面的提示是不是会得出什么结果。我想会的，不过至今我还未能作到这点。

对某些练习的说明

3．比如说$n=5$时就象图29所示那样描5个矩形。

4．令数b＝交于一个顶点的区域的平均数。那么证明就和定理2类似了。

5. 只有三角形ABC提供奇数个卵石。其它三

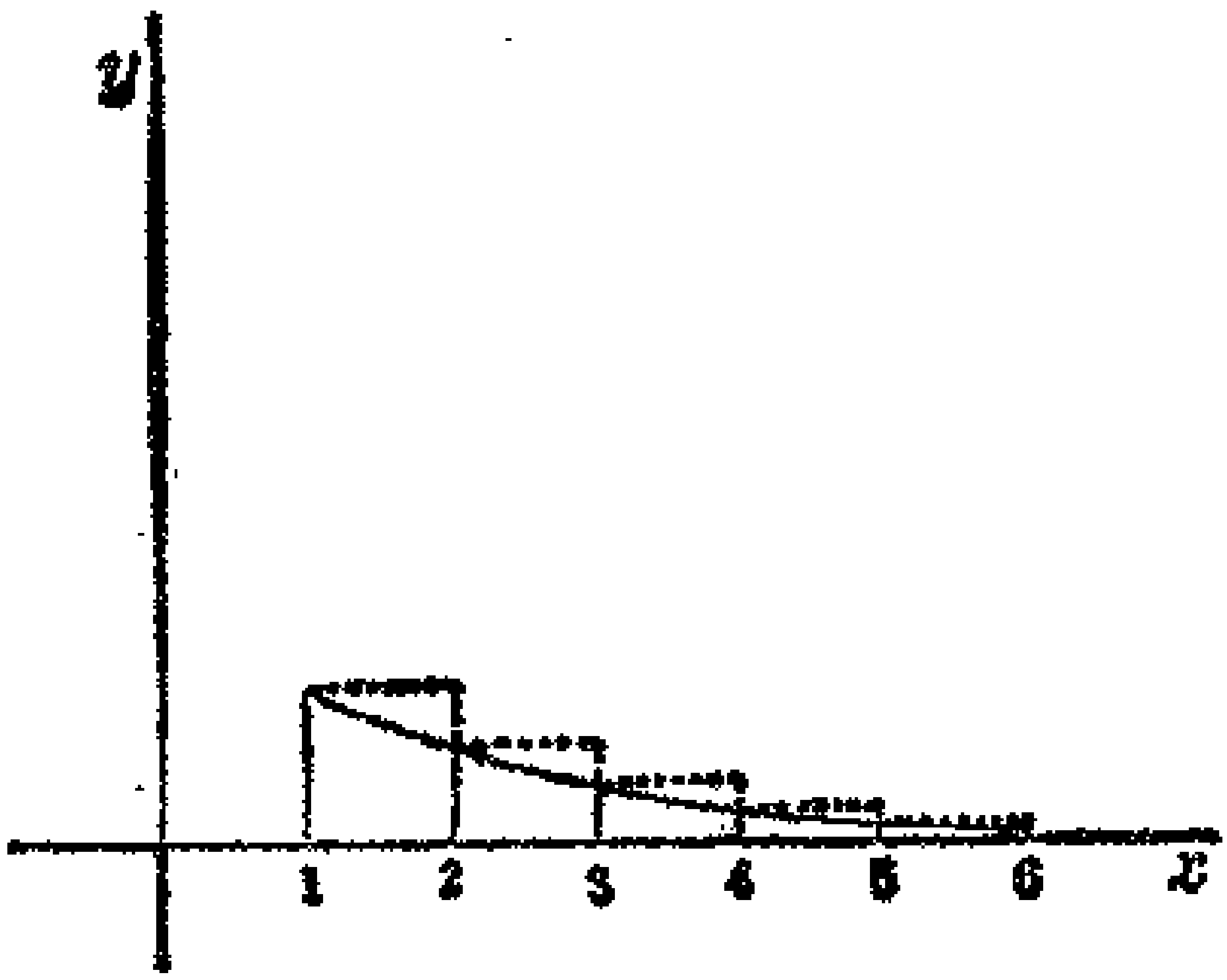


图29

角形，如AAB，BBB，AAC等，只提供0个或2个卵石，因而是偶数个。然后用两种方法数卵石的数目。

6. 证明类似于定理3的证明，但有几点细节不同。

8. 证明与定理3的证明类似，但有更多的具体细节不同。

9. 谁也不知道。但不难构造出一个方阵来，其中没有一个横截集合有n个相异数。

10. n大时，约为$(1-1/e)n \approx 0.63n$. 见[6]。

12. 的确成立，只消把后m列移向前去，无需重新证明。

13. 见练习6.

14. 与定理4的证明相似。

15 (a)这就是$m=n/2$, $q=2$时的定理4。可以设A对应于前$n/2$列。

(b) 这等价于(a).

18. (a)有4行在它的前2列中的部分至少有6个相异元素。

(b)还不曾有人进行过改进工作。

20. 首先交换积分顺序(付比尼原理)。

21. (a)用26除一个整数时，余数可能是什么？然后应用狄里克雷原理.

(b)是(a)的另一种说法。

22. 与十进展式的关系在大多数数论书中都有讨论。自己努力去证一下，不难。

23. (b)提示：把每一个都写成$2^a b$的形状，这儿a是非负整数，b是奇数。

24. 全部根之和与a_1有何关系？全部根之和与a_n又有何关系？

26. 按某个顺序给全部区域编上号：$C_1, C_2, C_3, \cdots$.考虑已给地图的那些包含区域C_1的全部有限子图。这种图有无限多个。由于只有5种颜色，所以有一种颜色要在这无限多个图中用来给C_1涂色。就用这个颜色来涂C_1。从这以后只考虑那些有限图。其中有无穷多个(有限)图，在那里面区域

C_2具有相同的颜色。就用此色涂 C_2。如法继续作下去。

27. 沿边界求曲率半径的积分。

参 考 文 献

1. N. Abramson, Information Theory and Coding, McGraw Hill, New York, 1963. (See pp. 169—173, etc., where the principle of randomness is discussed.)

2. L. D. Baumert, et al., A combinatorial packing problem, Computers in Algebra and Number Theory, SIAM/AMS Proceedings, 4(1971). (This uses the Fubini and Dirichlet principles to establish the existence of large packings by 2-by-2 squares on the surface of a torus, see p. 99.)

3. C. A. Rogers, A note on coverings, Mathematika, 4(1957)1-6. (This uses the Fubini and Dirichlet principles to establish the existence of thin coverings of space by spheres.)

4. ——, probabilistic and combinatorial methods in the study of the geometry

of Euclidean spaces, Proc. Int. Cong. Math, Vancouver, B. C. , 1974, col. 1,497-500. (A sketch of the method used in [3].)

5. S. K. Stein, Mathematics, the Man-made Universe, Freeman, San Francisco, 1976. (Exercises 44-56, pp. 269-270, present various applications of the pigeonhole principle. This is an elementary text.)

6. ——, Transversals of Latin squares and their generalizations, Pacific J. Math. 59(1975)567-575. (This is related to Theorem3.)

7. ——, Two combinatorial covering theorems, J. Combinatorial Theory, 16(1974) 391-397. (Some generalizations of Theorem 4.)

8. ——, The symmetry function in a convex body, Pacific J. Math., 6(1956)145-148. (This uses the Fubini and Dirichlet principles to establish the existence of a large symmetric set in any convex set.)

9. H. S. Stone, Discrete Mathematical Structures and Their Applications, Scien-

ce Research Associates, Chicago, 1973. (See pp. 118-119 for the use of the Fubini principle in a group-theoretic counting device, *Burnside's theorem.*)

五、概率奇事

R.洪斯贝格尔

1.埃夫伦骰子 在《科学美国人》1970年12月号上，马丁·加德纳在他所主持的“数学游戏”专栏里介绍了四颗不寻常的骰子。这副骰子是斯坦福大学统计学家布拉德雷·埃夫伦的发现。这四颗骰子A，B，C，D示于图30。假设我和你象下面介绍的规则来玩这副骰子：你拿一颗骰子，我拿另一颗；我们各自掷自己拿的骰子，谁得的点数大谁就赢。

如果骰子A与B来比，显然A得的点数大些的概率是2/3：

$\frac{2}{3}$的次数A的4胜B的3，

$\frac{1}{3}$的次数A的0败于B的3。

同理，B的点数比C的点数大的概率也是2/3：

$\frac{2}{3}$的次数B的3胜C的2，

$\frac{1}{3}$的次数B的3负于C的6。

而C又以$\frac{2}{3}$的概率战胜D：

$\frac{1}{3}$的次数C的6点取胜，

余下的$\frac{2}{3}$次数的一半C的2点胜D的1点，

因此总共$1/3+1/2(2/3)=2/3$的次数C获胜。

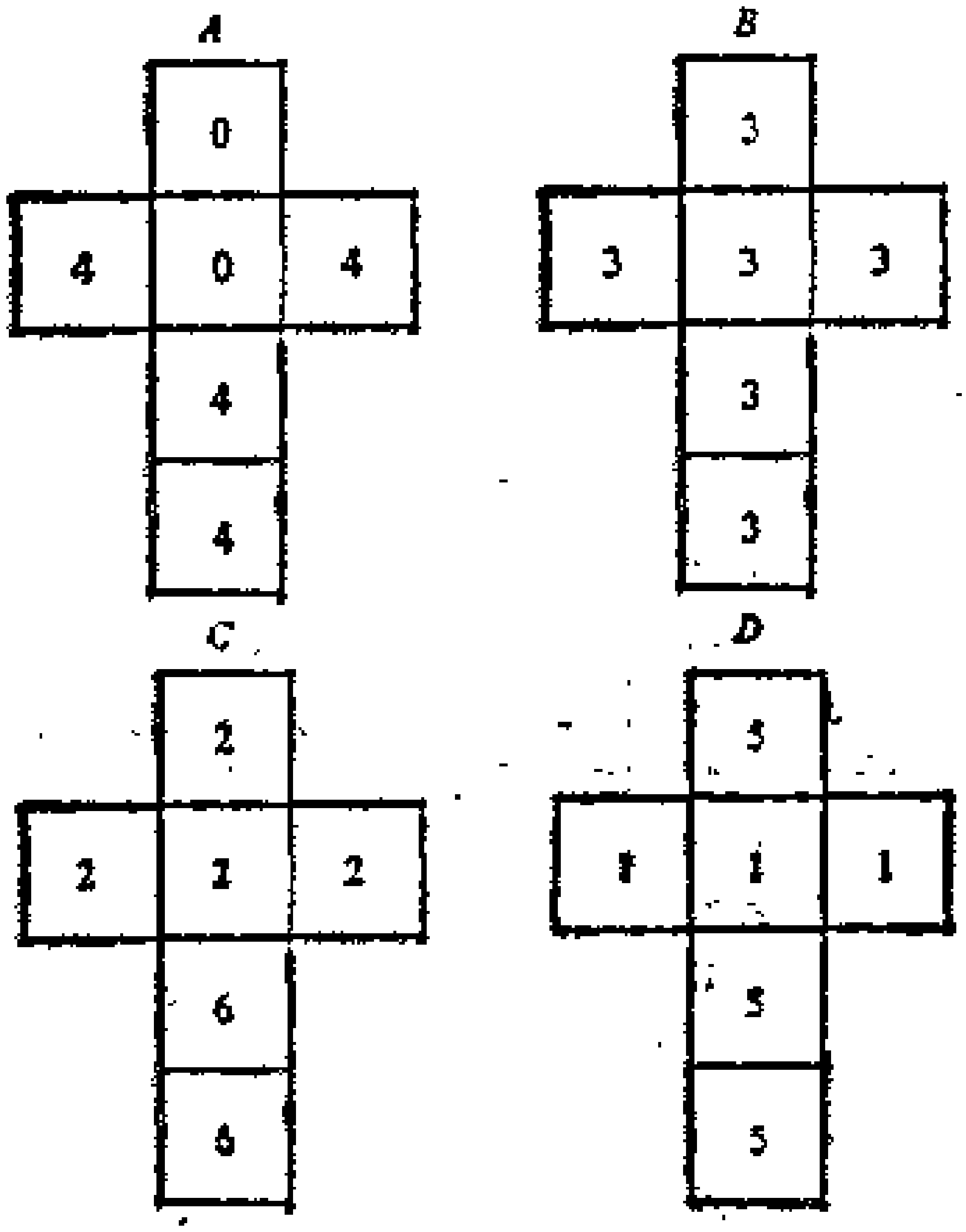


图30

因此，我们得到：

A与B比，A胜的次数可能是B的两倍；

B与C比，B胜的次数可能是C的两倍；

C与D比，C胜的次数可能是D的两倍。

这样一来，你会推测A胜D的情况．你可能 认为D胜A的概率理所当然的是$(1/3)^3=1/27$，然而，当

我们实际去算后，使人大为惊奇，它竟是2/3：

$\frac{1}{2}$的次数D的5点取胜，

余下的一半的$\frac{1}{3}$次D的1胜A的0，

总起来是$1/2+(1/3)(1/2)=2/3$的次数D获胜。

虽然可能看来不能令人相信，但这些骰子的確是非可递的。显然我们的游戏是“君子 游 戏”，因为你让你的对手先掷骰子。

图31和图32表示另外两副埃夫伦骰子。在第二

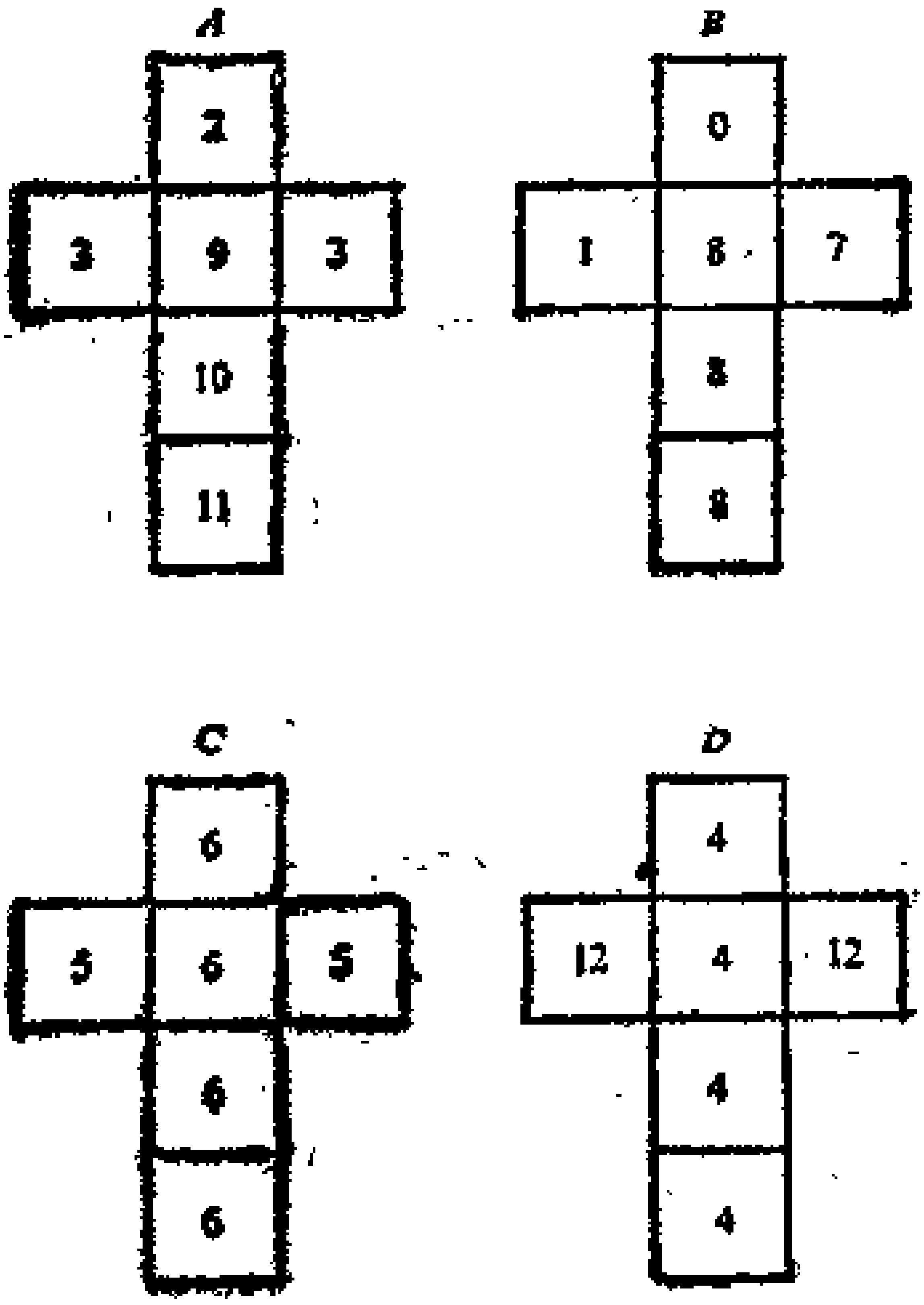


图31

副骰子里每对骰子可能有两个连结方式（比如我们的考虑中，就已经这样组配过：（A，B），（B，C）等等）在组配以后我们再来掷骰子。这就把这副骰子的一对对骰子的常数相对概率从2/3＝24/36减小到22/34。

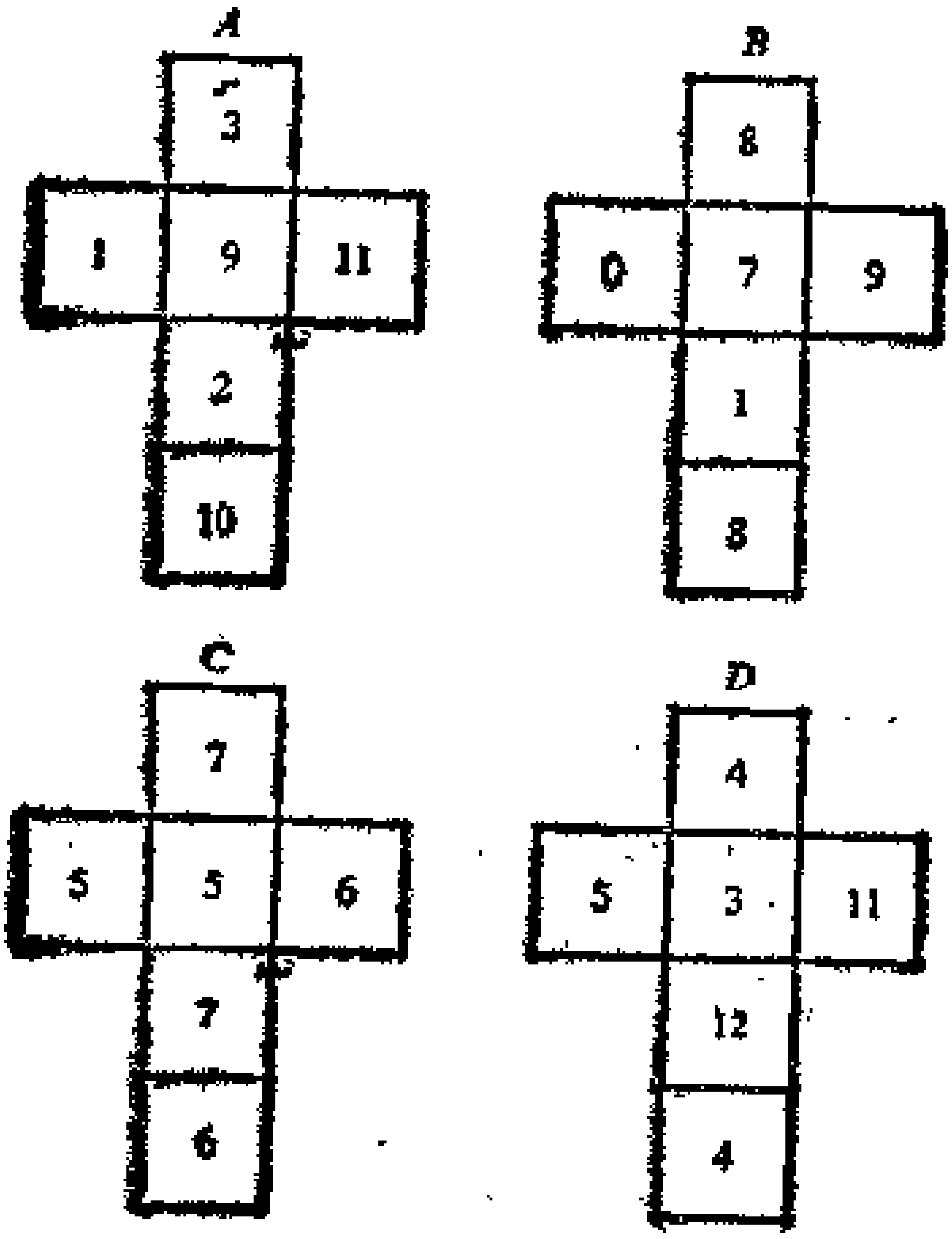


图32

2.血清 有一种疾病迅猛地在全国蔓延，得病的人一半死去，一半自愈。为了制止疾病蔓延制订了应急计划，迅速地生产了两种血清A和B。你可以料

想得到，几乎没有时间对这两种新血清进行试验。3位注射过血清A的病人都好了，8个注射血清B的病人有7个痊愈。你恰好去求医，听到医生说："你被传染了！"如果你不接受治疗，你活的可能性是50对50。但是这两种血清效果却优于这种情况。于是在手头证据尚不充足可靠的情况下，你愿意用哪种血清？

虽然在这个例子中的数目都非常小，但这种性质的问题却是统计学家最关心的问题之一。3例中有百分之百成功，8例中有7例成功，这二者如何能作比较？尽管八分之七是更大的样本上的相当不错的比率了，可是百分之百的记录使人心里觉得多安稳啊！在这类事情上除了述诸直觉外看来别无良法。不过有一个很巧妙的思想，即所谓的"虚假设"使我们可能对分析情况进行测度，从而对这种例子的情况作出明确的取舍。

血清A既无益也无害时接受血清的3例都痊愈的可能性有多大？由于未接受治疗的病人有50比50的痊愈的机会，不难看出这个概率是

$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}.$$

但是，如果血清B不生效时，8例中有7例痊愈的概率是

$$8\left(\frac{1}{2}\cdot\frac{1}{2}\cdots\frac{1}{2}\right)=8\left(\frac{1}{2}\right)^8=\frac{1}{32}$$

(这里的因子8表示有8种不同可能的受害者，死亡的概率也是$\frac{1}{2}$。因此已知的证据向人们表明血清A无效果的可能性是血清B无效果的可能性的4倍。据此显然应选用血清B。

3.校对员 《美国数学月刊》1976年第42页上德高望众的乔治·波里雅介绍了下列问题：

> 在校对一份手稿时A发现a个错误，B发现b个错误，其中c个错误也曾由A发现过。他们两人漏掉多少错误？

人们立刻作出的反应肯定是，漏看的错误数目可以是任意数。他们发现的应校正之处与他们没有发现的应校正之处有什么关系？当然不可否认未检查出的错误的数目可能是任意数。诚然在他们的成绩与缺点之间并无数学上确定的联系，然而这些事物的确象是同一个银币的不同两面。这种情况会使我们比仅仅瞎猜一阵好一点么？事实上我们通常的处境是可以有利一些的。

假设手稿上的错误总共有m个。发现a个错误时A就表明他的效率达到了a/m的水平。但是A还发现了B注意到的b个错误之中的c个错误，于是第二个数c/b又表现了他的效率是什么水平。虽然我们人人都是不稳定的，但是如果假设A工作时前后

一贯稳定也不算是瞎说．所以我们对A的评价就确定了我们对方程

$$\frac{a}{m}=\frac{c}{b}$$

及其引出的结论的相信程度．如果人们都很自觉尽责的话，我们满可以相信

$$m=\frac{ab}{c},$$

以及A、B二人漏校的错误数目

$$=m-a-b+c=\frac{ab}{c}-a-b+c$$

$$=\frac{ab-ac-bc+c^2}{c}=\frac{(a-c)(b-c)}{c},$$

这可由已知数量计算出来．

当$a=200$，$b=220$，$c=175$时，我们算出

$$\frac{(a-c)(b-c)}{c}=\frac{25\cdot 45}{175}=6.4.$$

由此可知校对大约漏了7处错误．

4.轮盘赌　琼斯先生带着20美元在拉斯韦加斯赌城去赌轮盘赌．他想赌到翻一番时或者赌到输光时就不赌了．他老是赌所谓的“平分”即黑或红，奇或偶．他考虑用下面的两个战术来赌．

(i)把20美元全部押在红色上，要么带着40美元要么分文全无就立即回家；

(ii)每次押1美元，一直赌到有40美元为止或

者赌到输光为止。

请问哪种赌博战术好些？

轮盘上除了有1至36个数外还有0和双零00。36个数中有18个是黑色有18个是红色，而0和00既不黑也不红，也不算成奇也不算是偶。所以0和00使赌场老板处于有利地位。赌“平分”的人赢钱的概率实际上只有18/38而不是1/2，尽管摆赌的老板是按押赌的人赢了时付给赢家双倍赌注的规则付钱。所以照第一种策略赌博，即一次押上20美元，赌赢的概率是18/38，约百分之47.4。

经过一番分析可以推导出按第二种策略押赌时赢钱的概率是

$$\frac{1-\left(\frac{20}{18}\right)^{20}}{1-\left(\frac{20}{18}\right)^{40}}.$$

我们关心的不是如何推导出这个式子，而是关心它的数值（有关这个问题的讨论请看弗雷德利克·莫斯特勒写的迷人的书《概率论难题50则》第54页以后各页；至于这问题中出现的差分方程的解法，则请参看萨缪尔·戈尔德贝格写的很好读的《差分方程》第三章，特别是第134页以下各页）。这个数真是小得惊人，才11%！换言之，一次押完赢的概率是4倍多。由此可见你赌得越久，摆赌的老板就越高兴。

5.掷钱币 假设人们一而再地掷一枚完好无缺的钱币，产生一个无限长的H与T的序列，H代表有人像的正面，T代表无人像的背面：

$H\ T\quad T\ H\quad T\ H\quad H\ H\quad T\ T\quad H\ T\cdots$,

顺便问一句，你觉得掷币一百次正好有50次得到正面的概率是多大？你可能以为是1/2左右，但却不是，而是百分之8还小一点（50处都是头像那一面的概率是

$$\binom{100}{50}\Big/2^{100},$$

其值约为0.0796）.

我们现在关心的是相邻的三元组在我们序列中出现的情况．这种三元组的数目是无止境的；他们出现在（1，2，3），（2，3，4），（3，4，5）等位置上．但三元组的类型只有八种（HHH，HHT，HTH，THH，HTT，THT，TTH，TTT）．如果我们集中注意于序列中的任意一个特别的三元组．比如说（5，6，7），那么这个三元组正好是8个中的指定的某一个组（例如HHT）的概率是1/8．在这一点上8个可能的三元组都是等价的．如果从序列开头依次考查三元组（即先看（1，2，3），再看（2，3，4），再看（3，4，5），一直依次看下去．你简直会吃惊得吓一跳，三元组TTH出现在三元组THT之前的可能性竟然会是出

现在THT之后的可能性的两倍。THH在HHT之先比在HHT之后的可能性会是三倍。在HTT，TTH，TTT的各种可能性中，HTT先出现是其它两个三元组先出现的可能性的六倍。类似的结果还有许多。

我们来证明上面提到的结论的前两个，先从第二个开始，即

> THH在HHT前出现的可能性是THH在HHT之后出现的可能性的三倍.

掷头两次，钱币币面可能是HH，HT，TH，或TT。现在考虑HH开头的情况：

$$HH\cdots\cdots.$$

显然此时THH不可能在HHT之前了：

> 为了得到THH，必须有一个T；但第一个T一出来，便组成了一个HHT。

但是在另外那三种情况，我们将会看到，HHT同样显然地不可能出现在THH之前，现在作一个相反的假设，假设HHT是序列

$$\cdots\cdots Z\ Y\ X\ \underline{H\ \ H\ T}\ .\ .\ .$$

中要出现的两个三元组的第一个，如果紧靠前的位置上的X是T的话，那么紧靠前的三元组就是$XHH=THH$，这就是THH先出现了。如果不是T，即$X=H$，那么我们就有

$$\cdots\cdots Z\ Y\ H\ \ H\ \ \underline{H\ T}\cdots\cdots.$$

如果现在相邻的先行元素Y是T，则THH在HHT之前。事实上为了不让THH先出现，所有的先行元素都应是H。不过在我们这里的三种开头—HT，TH，TT—的每一种中出现的一个T，就使这种情况不可能产生了。由于诸序列的1/4是以HH开头，其中3/4由HT，TH，TT之一开头，我们的结论便是：HHT有1/4的时候居于THH之前，有3/4的时候居于THH之后。这结果证明我们提出的性质是正确的。

但是很可能你会埋怨我们没有考虑一个序列不产生三元组THH与HHT的任何一个的可能性。这种可能性的确是有的，甚至会有无穷多个这种序列。然而尽管这样，不含全部8种可能的三元组的序列集合S只是序列总体中的一个无限小的部分，因此把它包含在H与T的序列的任一无限集合中或将其排斥在外，都不影响集合的"测度"（即S是测度为零的集合）。在以下的讨论中我们将会明白这一点。在证明我们的第一个结论（即TTH在THT之先的可能是TTH不在THT之前的可能的两倍）正确时，不必排除不确定序列的可能性我们就可以看到TTH在THT之前的概率是2/3，相反顺序出现的概率是1/3。所以二者之一出现在另一之前的概率是$2/3+1/3=1$，这意味着不确定的概率是0。

令A表示在以TT开头的序列中TTH在THT

前出现的概率。同样，令B，C，D分别表示以TH，HT，HH开头的各序列的相应的概率。我们注意到，这些概率同样考虑了确定序列与不确定序列。此时掷第三次钱币，就是用H或T来延长我们的序列，得出的三元组列在下面的表中。如果第一个三元组解决不了手头的问题，那么就把第一次掷钱币的结果忘掉不管算了，因为只有掷出的后两个结果才对构成序列中的下一个三元组有作用。考虑后两个结果(位置2、3上)，我们便能够确定每次延伸序列时TTH位在THT之前的概率。得出的结果列在下面的表中。

由于第三次掷钱币时出现一个H与出现一个T，可能性是完全一样大，所以每个延长的方式出现的概率都是1/2。例如，希望出现的事件$E=$（TTH在THT之先）发生的概率设为A，从TT开头，于是有一半的次数，序列按TTH延伸，由此得E的概率为1；另外一半的次数，序列为TTT，由此得E的概率又是A（只有最末两位字母以此概率出现）。由此得方程

$$A=\frac{1}{2}\cdot 1+\frac{1}{2}\cdot A.$$

另外情况下的相应结果列于表中。从得到的四个方程我们得出

$$A=1,\qquad B=\frac{1}{3},\qquad C=\frac{2}{3},\qquad D=\frac{2}{3}.$$

$$
\begin{array}{lll}
A: & TT\left\langle\begin{array}{l} TTH-1 \\ TTT-A \end{array}\right. & \Big\}\ A=\frac{1}{2}\cdot 1+\frac{1}{2}\cdot A,\ A=1; \\
B: & TH\left\langle\begin{array}{l} THH-D \\ THT-0 \end{array}\right. & \Big\}\ B=\frac{1}{2}\cdot D+\frac{1}{2}\cdot 0,\ D=2B; \\
C: & HT\left\langle\begin{array}{l} HTH-B \\ HTT-A \end{array}\right. & \Big\}\ C=\frac{1}{2}\cdot B+\frac{1}{2}\cdot A,\ 2C=A+B; \\
D: & HH\left\langle\begin{array}{l} HHH-D \\ HHT-C \end{array}\right. & \Big\}\ D=\frac{1}{2}\cdot D+\frac{1}{2}\cdot C,\ D=C,
\end{array}
$$

$A=1$,

$B=\frac{1}{3}$,

$C=D=\frac{2}{3}$.

而TT，TH，HT，HH这四种情况都有同样的可能性，这就意味着

E（即TTH居THT之前）的概率

$$=\frac{1}{4}A+\frac{1}{4}B+\frac{1}{4}C+\frac{1}{4}D$$

$$=\frac{1}{4}\left(\frac{8}{3}\right)=\frac{2}{3}.$$

用同样方法可以求出THT居于TTH之前的概率是1/3，这说明我们提出的第一个结论是正确的。关于第三个结论的正确性，留给读者证明以资练习。

6.最大数 在这最后一段里我们考虑两个形影相随的问题，这两个问题很出名。在莫斯特勒的《概率论难题50则》中也讨论了这两个问题。

（a）假设在十亿张纸条上写上十亿个不同的数，然后把纸条混合在一起放在一顶(大)帽子中。从帽子里你可以每次取一张纸条出来，你的任务就是找出写有最大数的那张纸条。关于那些数字的情况只给你说了它们是不同的数，但是关于它们的大小你却一无所知，所以取出的一些数一点也没有暗示待取各数的情况。难就难在你必须一抽出纸条就立即宣布你找到了最大的那个数；一旦取出一张纸条，你就再也不准说前一张纸条上有最大数（要不然你挑选之前只需把所有纸条浏览一遍就行了）。

纸条这么多，看来任务是完不成的了。你可以绝对相信，的确有一种策略存在，用它完成这项任务的概率比1/3还大。

为了好分析情况我们假定帽子里有n张纸条。抽取纸条就等于把纸条排了顺序，所求最大数X位于第i个位置的概率至少也是$1/n$。这里$i=1$，2，…，n，所以不管你抽出哪一张纸条，你赢的概率至少也是$1/n$。

不消说你绝不会挑选一个没有你已经抽过的数大的数.如果一个数连你已选的最大数都不是，那么它就不可能是全体纸条上的最大那个数了。所以在

多次抽取时，挑选的进行是自动地在继续。依照莫斯特勒的办法，把要选的数的最大的那一个称为“候选数”。只有在有候选数的情况才需要真正作出决断。第一个数永远是候选数，若不宣布它是最大数，就有危险漏掉所希望的最大 $1/n$ 的次数。但是如果总是选第一个数，这种策略只有 $1/n$ 的可能取胜。前面说过，再差也不会比这个结局更差了，所以我们尽可以心安理得地去寻求别的好办法。

考虑这样一种策略：先略去一定个数（比如 s）的纸条，然后在剩下的纸条中选取第一个候选数。要这项策略成功，必须有下列两个条件：

(a) 最大数 X 必须出现在位置 i 上，而 $i>s$，

(b) 位置 $s+1$，$s+2$，$\cdots i-1$ 上的数都不是候选数。

我们暂且只注意第 s 处以后的一个固定位置 i（见图33）。实现 (a) 的概率显然是 $1/n$。现在令 N 表示出现在前 $i-1$ 个位置上的最大数. 若 N 在头 s 个数中，则抽出 $s+1$，$s+2$，$\cdots$，$i-1$ 一个也不会是候选数。发生这种情况的概率是

$$\frac{s}{i-1},$$

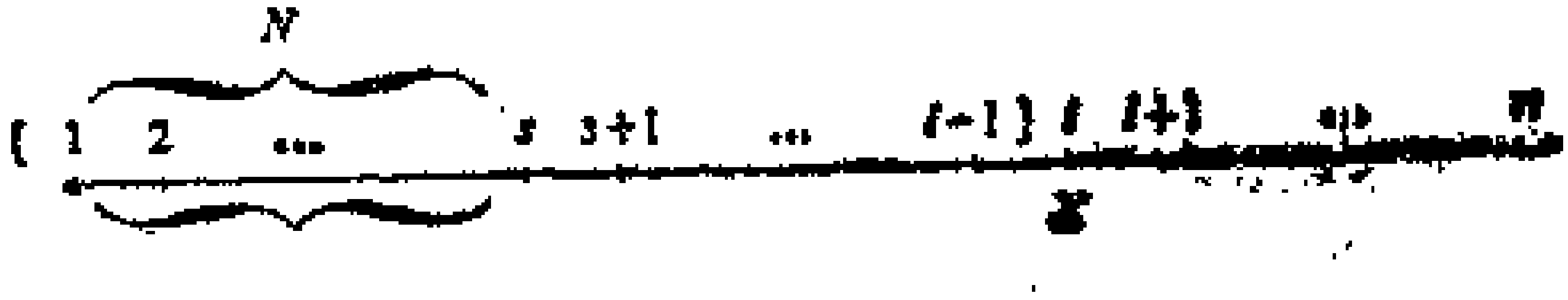


图33

因为它完全由所遇到的范围的相对大小来决定。于是我们这个策略在第i次抽取时成功的概率是$(1/n)(s/(i-1))$。由于这个策略在互斥情况$i=s+1, s+2, \cdots, n$的任何一种情况下都可能会取胜，所以用这项策略取胜的总概率是

$$P(s)=\sum_{i=s+1}^{n}\left(\frac{1}{n}\right)\left(\frac{s}{i-1}\right)$$

$$=\frac{s}{n}\left[\frac{1}{s}+\frac{1}{s+1}+\cdots+\frac{1}{n-1}\right]$$

当然$P(s)$的值视s之选法而定。我们的目的只是想说明存在一个s使得$P(s)>1/3$。

现在的情况是$n=$十亿$=10^9$。如果我们取$s=$368个百万$=368\cdot10^6$，即可得

$$P(s)=\frac{368\cdot10^6}{10^9}\left[\frac{1}{368000000}+\frac{1}{368000001}+\cdots+\frac{1}{999999999}\right]$$

$$=0.368\left[\frac{1}{368000000}+\frac{1}{368000001}+\cdots+\frac{1}{999999999}\right]$$

计算此式决非易事。不过我们现在关心的只是它的值而不是计算此值的详细过程。我确信方括号中和数的值很接近1（记住：调和级数是发散的）所以

$P(s)$大于0.36，显然比上面要求的1/3还大。

一般地可以证明s选得最好时总能得到一个$P(s)>1/e=0.3679\cdots$（e是自然对数的底），这个最佳s很接近n/e。n/e这个数用心算就常常可以得到相当准确的近似值，足以保证玩这种游戏取胜。如果玩的是一副普通的纸牌（比如梅花的2，3，…，幺点，方块的2，3，…，幺点，以及红心、黑桃的各点等等），要记的数是$s=19$($52/e=19.1\cdots$)。你的对家每次看一张牌，只把牌的背面亮给你，只保持最晚一个候选数（牌）。为了按照这个步骤进行，你丢掉19张摸上手的牌，选用这以后翻出的第一个候选数（牌）（你可以看到他换牌）。当然也可以不按最大数而按最小数来玩牌。事实上可以同时（用同一策略）玩两种。电子计算机或一张随机数表也是开始打这类牌的方便工具。

(*b*) 现在来考虑密切相关的另一个问题：如果人家告诉你帽子里纸条上的数是随机地从单位区间（0，1）中挑选出来的，请你从帽子中取出最大数。

在前面的问题中数字的绝对大小未起任何作用；不管纸条上的数字是什么，重要的是它的相对大小，这才决定它是否是候选数。讨论上下都有界的n个数时情况就不同了。由于这些数是随机选取的，人们完全有理由认为它们在区间范围内分布得

相当均匀。所以人们完全想要一个很靠近 1 的数，即使它是第一次抽出来的。不是候选数的数仍然不该选。就本题而言还必须考虑候选数的大小。在100张纸条中，第95次抽出的一个候选数一般说来是很有希望的数了，但若它是0.2，还是不宜冒险用它。

假设候选数是在第 $(n-r)$ 次抽出的，即后面还有r次可以抽。但是（0，1）中随机取一个数会$\leqslant x$的概率是x本身，这个数落在从x起头往下数的范围内的概率是 $1-x$（参看图 34）。另外请注意，一次抽数$\leqslant x$的概率与一次抽数$<x$的概率是相同的，一次抽数$\geqslant x$的与一次抽数$>x$的概率是一样的。玩这个游戏时，以后还有可能再抽x。万一它在更远以后出现，我们就把它当成候选数吧。因为愈往后边它愈有资格了，因为后边没有多少机会再提高它的资格了。要是最大数X碰巧出现不止一次的话，不管它那次出现我们选它都自然会赢的。所以剩下的全部r次抽数$\leqslant x$的概率是x^r，也就是采用手边的候选数x而赢的概率。

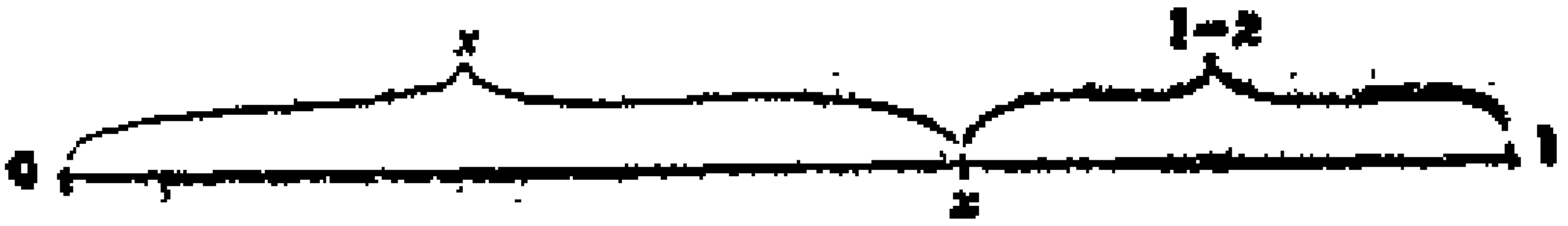


图34

拒绝手边的候选数x而采用下一个候选数（不管这候选数如何），这种时候赢的概率是多少我们现在就来解决。当然这会冒险，因为很可能后而根本没有候选数了。这时我们的策略成功的概率是零。然而x既然是候选数，它也就是抽至此处时所有数中的最大数了，因而在最后r张纸条中若有候选数存在的话，所求的最大数X必在其中。

现在假定在剩下的r个位置上可能找到i个$\geqslant x$的数。但是纸条上一个数$\geqslant x$或$<x$并不取决于抽取纸条的时间，由于前面曾假设纸条记录之数在单位区间上均匀分布，所以一次抽取$\geqslant x$的概率始终是$1-x$而一次抽取$<x$的概率始终是x。现在考虑，剩下r个位置上的某i个地方包含$\geqslant x$的数。所有i处有$\geqslant x$的数的概率是

$$(1-x)^i.$$

其余全部$r-i$处有$<x$的数的概率是

$$x^{r-i}.$$

于是在最后r中某特定数目i张纸条有$\geqslant x$的数而其余的纸条有$<x$的数的概率是

$$(1-x)^i x^{r-i}.$$

由于有$\binom{r}{i}$个不同方法区分$\geqslant x$的数的r个位置，所以后r个位置的恰好某i个含有$\geqslant x$的数的概率是

$$\binom{r}{i}(1-x)^i x^{r-i}.$$

现在如果已知正好有i个位置含有$\geqslant x$的数，那么我们的策略只能在X是所碰到的i个数的第一个时才会赢。由于X在这i个位置上的任一处出现的可能性都一样（X不可能是$r-i$个$<x$的数），所以x处于这种领头位置的概率只有$1/i$。因而在剩下的r次抽纸条时，对于i个候选数来说，我们策略赢的概率是

$$\frac{1}{i}\binom{r}{i}(1-x)^{i}x^{r-i}.$$

由于i可以取1，2，…，r的任何一个值，所以拒绝第$(n-r)$个位置上的一个候选数x而采用下一个候选数时赢的概率为

$$P=\sum_{i=1}^{r}\frac{1}{i}\cdot\binom{r}{i}\cdot(1-x)^{i}\cdot x^{r-i}=\binom{r}{1}(1-x)x^{r-1}$$

$$+\frac{1}{2}\binom{r}{2}(1-x)^{2}x^{r-2}+\cdots+\frac{1}{r}\binom{r}{r}(1-x)^{r}x^{0}.$$

r已知时这个概率P是x的函数。取$n-r$位置上的x而获胜的概率x^{r}也是x的函数。方程$P=x^{r}$的解$\bar{x}$给我们指出了候选数x的大小，但就所考虑的两条策略而言我们对此尽可以随意处之；因为$x=\bar{x}$时无论用哪个策略，我们取胜的机会都是一样的。例如$r=2$时方程$P=x^{2}$就是

$$2(1-x)x+\frac{1}{2}(1-x)^{2}=x^{2},$$

由此得$\bar{x}\approx0.6899$。而今情况是$x>\bar{x}$时有$P<x^r$，当$x<\bar{x}$时有$P>x^r$。所以如果我们在位置$n-2$上碰上的候选数$x<0.6899$，我们就继续往下进行去找下一个候选数，我们赢的机会就多些；若$x>0.6899$，我们就不去找下一个候选数而就用这个候选数还要好些；当$x=0.6899$时，两种策略成功的前景是一样的。

注意，$\bar{x}$这个值是不依赖于纸条总数n的，它只依赖于r。所以不管总共是一百张纸条还是一百万张纸条，如果倒数第三次抽出的$x<0.6899$就应该抛弃。尽管我们的分析很难说是绝对的，但它却使我们可以玩得聪明一点，因为它给了我们一个$\bar{x}$这个$\bar{x}$至少可以指示我们不要采用什么数（因为尽管继续下去也许不是最好，但总会好一些）。

但是我们的整个分析又面临一大困难，那就是从方程$P=x^r$求$\bar{x}$时计算量太大。莫斯特勒有一个很好的近似式

$$\bar{x}=\frac{r}{r+0.8043}$$

挽救了我们的方法。（要了解上述近似式的推导过程，请读莫斯特勒写的书第79页。）所以人们可以用手算得到十分可靠的结果，或者更轻松一点就用廉价的袖珍计算器来算。$r=1，2，\cdots，10$时算出的近似值列在下面的表中。

r	x
1	.5542
2	.7132
3	.7886
4	.8326
5	.8614
6	.8818
7	.8969
8	.9086
9	.9180
10	.9256

由于

$$\frac{r}{r+0.8043}=\frac{1}{1+0.8043/r}$$

随r增大而增大，所以可以当候选数的一个已知数出现得愈早就愈容易遭到拒绝。所以表上不是给出一些孤立的结果而是给出大量信息，例如，如果任意一个候选数<0.9086，那么都要放弃，后面还得抽8次或更多次。

和玩第一个游戏一样，若用一台电子计算机或一张随机数表，玩起来就很方便。在玩第二个游戏时为了玩得更富于竞争性，必须为x计算出很大的详表来。

练习

1．下面的3×3幻方名叫洛书，是最古的幻方

A:	6	1	8
B:	7	5	3
C:	2	9	4

图35

（图35）．设有九位棋手，按棋艺高低列为1，2，3，4，5，6，7，8，9等，按洛书的行分为三队A，B，C。问题是：哪个队最强？我们靠把各队队员实力相加也一点看不出孰优孰劣，因为每队总分都是15。实际下棋来确定优劣之类的办法也没有。所以安排进行循环赛。每次比赛共九局，让一个队的三名棋手中的每一个都与另一对阵队的三名棋手中的每一个对奕。试证A击败B，B击败C．显然A不如C。（最好检验一下）。

2.在一个大袋子里装有个数未知的普通白色乒乓球。抽出60个来涂红再装进袋中与白球混合。再任取出一百个球，发现其中有15个红球。问最初袋中有多少球？（关于这个问题的有趣讨论，请读费

勒写的《概率论基础及其应用》第一卷第三版45～46页）。

3. 在门口的一边有一个瓮A装有70只球，在另一边有一个瓮B装有30只球。每个经过此门的人任选一个瓮，从里面取出一只球放进另一个瓮中。假设有数量不限的人经过此门，试问A被取空的概率是多少？B被取空的概率是多少？A和B一个也不空的概率又是多少？

4. 一枚完好无缺的钱币被多次抛掷后产生出一个H与T（即人头像正面与无人头的反面）的序列。试证明邻接三元组HTT在序列中出现在TTH或TTT之前的可能性是后二者任一个出现在HTT之前的可能性的六倍。

5. 巡官只知道他正在追捕的贼是他从街对面僻巷监视的夜总会的客人中最高的那个。为了逮捕罪犯，巡官必须把他跟踪到家连人带脏一起捉获。客人一个一个地离开夜总会时要经过一个灯光明亮的地段，所以巡官可能判别人们的高矮。困难的是有40个人，但只有一个才是贼。试问巡官应该尾随哪一个人才最有可能跟对？

6. 两张红牌和两张黑牌洗后面朝下放成一行。随意地取两张牌，这两张牌颜色相同的概率是多少？

7. 10只白球和10只黑球应该如何分放在两个盒

子里，才能使人任意从一个盒中任取一个球出来得到白球的机会最好？（例如10个黑球放在一个盒子，10个白球放在另一个盒子中，得白球的概率是$\frac{1}{2}(1)+\frac{1}{2}(0)=\frac{1}{2}$。同理，每盒各放5个白球5个黑球概率也一样。但是高达0.73的概率却是办得到的。）

8.保管帽子的女郎给n个进剧场看戏的人看管帽子.不巧他把全部帽子完全弄混了.她不愿承认有错，看戏的人出戏院时她完全随意地把帽子递还给观众，希望他们中间有些人能拿回自己的帽子。如果$n=2$，至少一人取回的是自己的帽子的概率有多大？$n=3$，或4，或5时概率又是多大？（请注意一个明显的现象，就是这个概率随着n的增大而交替地忽大忽小。）

六、反常约分

R·P·波亚士

这条数学之弦弹出的是不正之音，但弦外之音却是严肃的。这则数学问题起源于64/16＝4/1（“约去了6”）这个现象。作为辩解之词，这对说“我用自己的方法得到了正确答案，不是吗？”这类话的学生很有用。

如果你是差等生，你对“64/16＝4/1”的反应多半是眼里充满了困惑莫解的神情。如果你稍稍能干些，你就会问还有没有这种现象的更多的例子，而我就可以拿出98/49，95/19，65/26（及其倒数）来作例子；你可能会继续问除了22/22这类例子外是否还有更多的例子。事实上已经知道再也没有更多的例子了（参见〔5〕，〔7〕，〔8〕）。这个问题好象为计算机天生的题目一样，因为计算机把所有二位数相除的商都搜索了，只用了0.15秒钟就把这4个例子搜索出来了。

我们现在想知道这是不是基为10的特有现象。举例来说，在十二进制中，我们的情况更好些呢还是更差些？把问题提得更一般化些，我们把整数

用基b表示，试问何时有

$$\frac{x,\ y}{z,\ x}=\frac{y}{z},$$

式中x，y，z是b进制下的单个数字，并且

$$\frac{x,\ y}{z,\ x}\ 的意思就是\ \frac{bx+y}{bz+x}.$$

这里不考虑倒数16/64等，因为我们可以倒起约它们。很显然，除了最平凡的情况外我们必须错开约，并且不能有位数等于0。

通过提一般化的问题，我们就把孤立的好奇事例带入了数论的一个虽属初等但却要认真对待的论题中了。原来这和一些很深入的问题有关系，而且还给计算机的使用开创了新的可能。

对于2进制中的情况显然不感兴趣，对3进制也不感兴趣，因为只有两个不为零的数字。以4为基时，一试就得到

$$\frac{3,\ 2}{1,\ 3}=2$$

但却难于从这么一点儿实例便进行推广。

为什么我们不叫计算机来做这件事呢？b很小时倒很好，但是b一大，分数便太多了，事实上有$(b-1)(b-2)(b-3)$个。核查一个分数要进行7步，每一步按耗时最少算也要用1微秒。$b=10$时，这倒微不足道，但$b=1000$时却要花费2小时左右（实际工作时费的时间还要长些）。寻

点基本理论作准备，我们就能工作得好些，我们要讲解一个原理，不过要先想想才开始算。

我们来求

$$\frac{bx+y}{bz+x}=\frac{y}{z}$$

的解，这里x，y，z都在1与$b-1$（包括$b-1$在内）之间，任何两个都不相等。

定理1．如果b是素数，方程无解。

用$z(bz+x)$乘方程得

$$b(x-y)z=y(x-z). \qquad (1)$$

b为素数蕴涵$b|y$（和平常在数论中一样是指"b整除y"）或$b|(x-z)$。但由于$1\leqslant y\leqslant b-1$，$1\leqslant|x-z|<b-1$，因此所说的整除一个也不可能。

定理2．b的每一个异于b与1的因子，都对应着一个解，当b－1是素数时解还是唯一的。

重新排列（1）我们得

$$yz(b-1)=x(bz-y).$$

假设$b-1$是素数，则$b-1$整除x或$bz-y$。在后一情况它也就整除$(b-1)z+z-y$，因而整除$z-y$；然而这是不可能的，因为$1\leqslant|z-y|<b-1$所以$b-1|x$，意即$x=b-1$。于是我们得$yz=bz-y$，因而y是z的倍数，令$y=kz$得$kz^2=bz-kz$，$y=kz=b-k$，因而k是b的因子．最后得到$x=b-1$，$y=b-k$，$z=(b/k)-1$。我们把这三者缩写成

一个符号：

$$(b-1,\ b-k,\ \frac{b}{k}-1).$$

直接代入即可证明，不管$b-1$是否是素数，这都是解。

所以解的个数至少是b的真因子的个数，比如$b=12$时为4，$b=60$时为10。我将称对应的解为平凡解；我们将会看到b和$b-1$都是复合数时恒有非平凡解。

我还不知道，求一般情况时的解的公式，但是却有一个算法，求解时比起一一核查b进制的二位整数要少很多步骤。

定理3．所有的解都是$x=mp$的形式，式中p是$b-1$的一个素因子，$1\leq m\leq r=(b-1)/p$；$y=x-k(1\leq k<x)$；$z=xy/(y+kb)$．只要z是整数这就是解。

定理3的证明很烦琐，我们把它放在本文的附录1中。

根据定理3可以制定一个很实用的算法，我给这个算法编制了计算机程序，而且算出了直到$b=100$的解，作为一个表。但是一看算出的表却发现许多原先未曾料到的情况，这些情况从已有的公式中不是一眼便立即看得出来的。有些情况事实上还要求对x，y，z作进一步的限制，所以还可以进一步减少计算量。

首先我们证明b与$b-1$为复合数时一定有非平凡解（满足$x<b-1$的解）。事实上若b是偶数，那么，由于p与r是$b-1$的因于因而p与r都是奇数。很容易验证$x=p(r+1)/2$，$y=b/2$，$z=(r+1)/2$是一个解。因$b-1>p$，它是非平凡解。所以

$$x=(pr+p)/2=((b-1)+p)/2<b-1.$$

如果$b-1$是奇平方，$p=r$，那末交换p与r还可得到第二个非平凡解。

如果b是奇数，s是b的真因子，则平凡解（$b-1$，$b-s$，$(b/s)-1$）的分量都是偶数，把它们都除以2就得到一个非平凡解。例如（8，6，2）在9进制中是一个解，所以（4，3，1）也是解。

事实上以9为基时没有别的解了。对于其它的奇数基b，若b与$b-1$为复合数，则有非平凡解存在，它们不能通过平凡解各位数除以公因子而得到。这一情况看来是真实的，不过我没有它的证明。

从解表中获得的最显然的结论之一是：y绝不是很小的数。

定理4．对一切解均有$y\geq(x+1)/2$**（因而有**$k\leq(x-1)/2$）。

如果用x，k，b来表示z，由于$z\geq1$我们得

$\therefore \dfrac{x^2-kx}{(b-1)k+x}\geqslant 1$，　$x^2-kx\geqslant(b-1)k+x$，

$$k\leqslant\frac{x(x-1)}{(b-1)+x}=\frac{x-1}{1+(b-1)/x}\leqslant\frac{x-1}{2},$$

因为$x\leqslant b-1$，所以$2\leqslant 1+(b-1)/x$。

我们还注意到x也绝不是很小的数。

定理5. 对一切解均有$x\geqslant 1+b^{\frac{1}{2}}$。

我们有$x=y+k>y$。我们刚才得到

$$\frac{x(x-1)}{(b-1)+x}\geqslant k,$$

所以由于$k\geqslant 1$，得到$x^2-x\geqslant b-1+x$，由此得$x\geqslant 1+b^{\frac{1}{2}}$。

b为平方时，比如$b=q^2$，一个解是$(q+1,\ q,\ 1)$，此处有$x=1+b^{\frac{1}{2}}$，所以定理5的等式部分成立。

现在我们可以把定理3的算法加以修正：$x=mp$，式中p是$b-1$的素因子，且$(1+b^{\frac{1}{2}})/p\leqslant m\leqslant(b-1)/p$；$y=x-k$，其中$(x-1)/2\geqslant k>1$；$z=xy/(y+kb)$。这就使算法变得很短，对于$b$很小的情况用手算就能求出解来。例如$b=10$时$p$必须等于3，$x=3m\geqslant 5$，所以$x=9$或6。$x=9$时得平凡解(9，8，4)和(9，5，1)。$x=6$时$y\geqslant 4$，所以$y=4$或5，于是得出另外那两个解：(6，4，1)和(6，5，2)。

事实上这个算法短小得可以（用计算机）分析$b=1000$的情况而得知有126个解，例如

$$\frac{629,\ 625}{85,\ 629}=\frac{625}{85}\text{和}\frac{501,\ 334}{1,\ 501}=334.$$

我们又注意到z始终是大大地小于y。

定理6. 对一切解均有 $z\leq y/2$。

证明留作练习。

长期实验还观察到，解常常是成对出现的，对每种基来说个数是偶数，但除去4，16，36，64与100等基。这些数据强有力地支持了下列定理。

定理7. **对于已知的b，如果b不是偶平方数，解的个数是偶数。**

事实上我们将会看到解是成对出现的，z与k互相交换，但要除掉b是偶平方数的情况，因为这时有$b^{\frac{1}{2}}-1$个解，$z=k$.

为了明白交换z与k总是能作出一个解，我们可以计算出，如果$(x, x-k, z)$是解，则条件$z=xy(y+kb)$（定理3）等价于$b=((x/k)-1)((x/2)-1)$。由于此式关于z，k对称，所以交换z，k又得一解$(x,\ x-z,\ k)$。这是一个不同的解，除非$z=k$，而这意味着$b=y^2/k^2$，于是b是平方数。因为$(b^{\frac{1}{2}}+1,\ b^{\frac{1}{2}},\ 1)$是解，故乘以$t$也是解，$t=2,\ 3,\ \cdots,\ b^{\frac{1}{2}}-1$，所以有$b^{\frac{1}{2}}-1$个不成对的解。这个数目是奇数的充要条件是$b$为偶数。

我们还看到，对给定的b通常至少有一个解合乎$z=1$，第一个例外出现在$b=35$时。事实上下面定理表明有无穷多个例外基。

定理8。当$b=2q+1$，q为奇素数，b为不能被3整除的复合数时，则没有合乎$z=1$的解。这种b值有无穷多个。

定理8的b的头几个值是35，95，119，143。定理中所说的b看来很可能是唯一没有合乎$z=1$的解的b了。头几个b值还提示我们，b总是两个素数之积，但$275=2\cdot137+1=5^2\cdot11$是个反例。

定理8的第一部分在本文附录1中证明。为证明b值有无穷多，我们注意狄里克雷定理〔6〕，这个定理说：具有一个以上素数的算术级数实际上就有无穷多个素数，根据这个道理形如$q=15m+2$的素数有无穷多，于是$2q+1=30m+5$可被5整除（因而是复合数）但不被3整数。

唯一只有一个解的基是$b=4$。

定理9。$b=9$时或者$b-1$是素数且b只有两个素因子时(其中一个必须为2)，只有两个解。

证明留作练习。

由此可知，为了要只有两个解，我们必须有$b=2q$，其中q是奇素数，或者$b=8$或9。是否有无穷多b值取决于是否有无穷多个素数$2q-1$，此处q是素数。这是尚未解决的、或许还非常困难的问

题.

绝不是正好3个解（练习3）．$b=10$，12，以及许多值时有4个解。可能出现4个平凡解的情况，但令人感兴趣的是两个平凡解和两个非平凡解的情况。当$b>9$时就会出现这种情况，具体就是$b=q^2+1=2t$时，式中q为奇素数，t为素数；$b=10$就是这样。b有无穷多个值的情况又是一个未解决的问题，这些b相对说来很不常见；头几个是10，26，122，362，而且猜出其格式是不容易的。

我们可能还想知道哪些基的解不是一般化的多。最可能的基看来是这样的一些基b，对它们来说b是平方数，$b-1$有很多因子。例如$b=36$时有21个解；$b=100$时有37个解；$b=441$时有130解；$b=1225$时有236个解，（近一个记录是）$b=3025$时有410个解。

下表列有头几个基和基为100时的全部解。

基4　1个解

x	y	z
3	2	1

基6　2个解

x	y	z
5	4	2
5	3	1

基8　2个解

x	y	z
7	6	3
7	4	1

基9　2个解

x	y	z
4	3	

8	9	2

基10　4个解

x	y	z
6	5	2
6	4	1
9	8	4
9	5	1

基12　4个解

x	y	z
11	10	5
11	9	3
11	8	2
11	6	1

基14　2个解

x	y	z
13	12	6
13	7	1

基15　6个解

x	y	z
12	10	3
12	9	2
14	12	4
14	10	2
7	6	2
7	5	1

基16　7个解

x	y	z
9	8	3
9	6	1
15	14	7
15	12	3
15	8	1
5	4	1
10	8	2

基18　4个解

x	y	z
17	16	8
17	15	5
17	12	2
17	9	1

基20　4个解

x	y	z
19	18	9
19	16	4
19	15	3
19	10	1

基21　10个解

x	y	z
8	7	2
8	6	1
10	9	3
10	7	1
16	14	4
16	12	2
20	18	6
20	14	2
15	14	6
15	9	1

基22　6个解

x	y	z
12	11	4
12	8	1
21	20	10
21	11	1
14	12	3
14	11	2

基24　6个解

x	y	z
23	22	11
23	21	7
23	20	5
23	18	3
23	16	2
23	12	1

基25　6个解

x	y	z
6	5	1
12	10	2
16	15	6
16	10	1
18	15	3
24	20	4

基26　4个解

x	y	z
15	13	3
15	12	2
25	24	12
25	13	1

基27　6个解

x	y	z

20	18	5
20	15	2
26	24	8
26	18	2
13	12	4
13	9	1

基28　10个解

x	y	z
9	8	2
9	7	1
15	14	5
15	10	1
18	16	4
18	14	2
27	26	13
27	24	6
27	21	3
27	14	1

基30　6 个解

x	y	z
29	28	14
29	27	9
29	25	5
29	24	4
29	20	2
29	15	1

基32　4 个解

x	y	z
31	30	15
31	28	7
31	24	3
31	16	1

基34　6 个解

x	y	z
18	17	6
18	12	1
33	32	16
33	17	1
22	20	5
22	17	2

基35　6 个解

x	y	z
24	21	4
24	20	3
34	30	6
34	28	4

17	15	3
17	14	2

基36　21个解

x	y	z
10	9	2
10	8	1
20	18	4
20	16	2
25	24	10
25	15	1
30	27	6
30	24	3
35	34	17
35	33	11
35	32	8
35	30	5
35	27	3
35	24	2
35	18	1
7	6	1
14	12	2
21	18	3
28	27	12
28	24	4
28	16	1

基38　2个解

x	y	z
37	36	18
37	19	1

基39　6个解

x	y	z
28	26	7
28	21	2
38	36	12
38	26	2
19	18	6
19	13	1

基100　37个解

x	y	z
27	25	3
27	24	2
33	32	8
33	30	3
33	25	1
42	40	7
42	35	2

51	50	17	99	98	49
51	34	1	99	96	24
54	50	6	99	96	19
54	48	4	99	90	9
66	65	26	99	80	4
66	64	16	99	75	3
66	60	6	99	50	1
66	50	2	11	10	1
66	40	1	22	20	2
81	80	36	44	40	4
81	76	9	55	50	5
81	72	6	77	75	21
81	45	1	77	70	7
84	80	14	77	56	2
84	70	4	88	80	8

附录1.定理3的证明. 我们必须求出$yz(b-1)=x(bz-y)$的整数解x, y,z来。取$b-1$的一个素因子p,则$p|x$或$p|bz-y$。我们证明在第二种情况时x为$b-1$的一个素因子所整除;我们称此因子为p,然后继续下去. 事实上我们有$bz-y=(b-1)z$

$+z-y$，所以若$p|bz-y$我们有$p|y-z$，于是$y-z$为p的倍数，令$y-z=qp$。由于$r=(b-1)/p$，$|y-z|<b-1$，我们有$q<r$。但是

$$yz=x\frac{bz-y}{b-1}=\frac{x(rz-q)p}{b-1}=x\left(z-\frac{q}{r}\right),$$

而此式只有在x与r有公因子t时才可能是整数。由于$r|b-1$，我们有$t|b-1$，所以x可被$b-1$的一个因子整除，这个因子我们可取成素数。

我们现在假设p为素数，$p|b-1$，$p|x$。于是$x=mp$，$b-1=rp$，其中$m\leqslant r$，因为$x\leqslant b-1$。所以我们有

$$(bz-y)m=yzr, \tag{2}$$

故$z|my$，设$my=sz$，$s>0$。由此得

$$mbz-sz=\frac{sz}{m}rz, \quad mb=s\left(1+\frac{rz}{m}\right),$$

$$m^2b=s(m+rz),$$

所以$s|m^2b$。现在我们就可以用s来表达x,y,z了：$z=my/s$，$y=(mb-s)/r=x+(m-s)/r$。这说明$r|m-s$。而今$m\leqslant r$，而$s>0$，所以我们得$m-s\leqslant 0$。但是$m=s$意味着$x=y$，这是我们已经排除了的情况。所以$m<s$，故$z<y$。另外，若我们写成$(m-s)/r=-k$，我们就有$k>0$所以$y<x$。

由(2)得

$$ryz=mbz-my, \quad z=\frac{my}{mb-ry}$$

$$= \frac{xy}{xb-(b-1)y} = \frac{xy}{y+kb}$$

于是当且仅当z是整数时有一个解是$x=mp$，$y=x-k$，$z=xy/(y+kb)$。如果我们对$b-1$的所有素因子都这样来构作x，y，z，我们就得到全部解。

定理8的证明．令$b=2q+1$，q为奇素数，假设有一个解$(x, y, 1)$。于是

$$\frac{m^2p-mk}{m+rk}=1,$$

式中（由于$b-1$仅有的因子是2和q）$p=2$，$r=q$或$p=q$，$r=2$。

由于z，k与k，z同时产生解，我们一定有

$$k=\frac{m^2p-m}{r+m}, \qquad 当z=1时。$$

情况（a）．

$$p=2, \quad r=q, \qquad k=\frac{2m^2-m}{q+m}$$

$$=m\frac{2m-1}{m+q}.$$

由于$1 \leqslant m \leqslant r=q$，且$q$为素数，所以$m+q$对$m$是素的，除非$m=q$；但此时$k=(2q-1)/2$已不是整数了。故情况（$a$）不出现。

情况（b）．

$$p=q, \quad r=2, \qquad k=\frac{m^2q-m}{m+2}$$

$$=\frac{m(mq-1)}{m+2},$$

式中m可以是1，也可以是2。

若$m=2$，则k不是整数。

若$m=1$，则$k=(q-1)/3$的充要条件是$q-1$为3的倍数，于是$2q+1$也是3的倍数。

附录2.

推广. 约去由2位以上的整数组成的分数的数字，至今尚无什么系统的理论。几个常常引用的例子（〔3〕，〔4〕，〔8〕）〔1〕据说是1935年由E·南内发现的；这就是下面的（E）、（F）、（G）。（在（G）中出现有1935的字样表明，这个例子的确很可能是在那一年或挨近那一年构造出的。）

另一个常引用的例子是

$$(A)\quad \frac{2(666)}{(666)5}=\frac{2(66)}{(66)5}=\frac{2}{5}.$$

（从这里起要约去的数字写在括号里。）几个不同类型的三位数例子如下：

$$(B)\quad \frac{(7)75}{21(7)},$$

$$(C)\quad \frac{1(2)(7)}{(7)6(2)},$$

$$(D)\quad \frac{5(9)4}{3(9)6}.$$

〔9〕和〔10〕中还收集了许多例子。(*B*)型例子很稀有；除了平凡例，同型例的倒数，以及(*A*)中把数字重复而得的数（参看下面定理10）之外，显然只有另外一个例（在十进制中）。〔3〕和〔10〕中给出的(*C*)类例也很稀少；其中某几例可以通过将数字重复的办法产生出更多的例子来，比如

$$\frac{1(66)(33)}{(33)2(66)}=\frac{1(6)(3)}{(3)2(6)}=\frac{1}{2}$$

但是象(*D*)的例子却极其普遍，已经晓得的就有好几百个了。另外一些例子见〔3〕，〔9〕，〔10〕。

南内的例子是：

$$(E)\quad \frac{3(5)44}{7(5)31},$$

$$(F)\quad \frac{143(18)5}{170(18)560},$$

$$(G)\quad \frac{425(1935)345}{912(1935)5185}.$$

（在我所见的例(*G*)被引用的地方，都把912错印成了918，尽管〔8〕、〔3〕都指出过这一误植；C.S.奥吉尔维告诉我说自从〔8〕发表后一些人已经看到勘误了。）

这些例子中有些看来很难，但难是难在少见。事实上我们将会看到类似于(*E*)，(*F*)，(*G*)

的例子还是很平常的。

我们先研究（A），更确切地说是研究（A）中分数的倒数。

定理10．如果以b为基对$\frac{x,\ y}{z,\ x}=\frac{y}{z}$，则也有$\frac{x,\ x,\ y}{z,\ x,\ x}=\frac{y}{z}$；反之，若$\frac{x,\ x,\ y}{z,\ x,\ x}=\frac{y}{z}$，则有$\frac{x,\ y}{z,\ x}=\frac{y}{z}$．

在第一条假设的前提下还有

$$\frac{x,\ x,\ x,\ y}{z,\ x,\ x,\ x}=\frac{y}{z},$$

而且还很可照此方式继续作下去。

定理10再加上刚才的补充说明，给了我们(A)中的例，以及十进制中三个解产生出的另外的例。因为以100和1000为基的例子可以当成是以10为基的例子来念，所以我们还有

$$\frac{(629629)625}{85(629629)},$$

这个例子就不大可能偶然碰得到了。

定理10的证明．已经知道

$$\frac{xb+y}{zb+x}=\frac{y}{z},$$

即

$$bxz+yz=byz+xy. \tag{3}$$

我们想证明

$$\frac{xb^2+xb+y}{zb^2+xb+x}=\frac{y}{z},\qquad (4)$$

即

$$b^2xz+bxz+yz=b^2yz+bxy+xy=b(byz+xy)+xy. \qquad (5)$$

把（5）左端的$bxz+yz$换成．（3）中的右端，同样把（5）右端的$byz+xy$也换掉．（5）化成

$$b^2xz+byz+xy=b(bxz+yz)+xy,$$

这是一个恒等式．

现在设（5）为已知；（5）本可写成

$$bxz+yz=b^2yz+bxy+xy-b^2xz,$$

于是有

$$bxz+yz-(byz+xy)$$

$$=b^2yz+bxy-b^2xz-byz$$

$$=b(byz+xy-bxz-yz)$$

$$=-b(bxz+yz-byz-xy).$$

如果不是$bxz+yz=byz+xy$，上式就表明$b=-1$．因此（3）成立．

证明（4）可以推出

$$\frac{xb^3+xb^2+xb+y}{zb^3+xb^2+xb+x}=\frac{y}{z}$$

的任务留给读者来完成．

接着我们来研究（D）类例子，这类例子中约去的是中间数位上的数字。例如

$$\frac{b^2p+by+q}{b^2r+by+s}=\frac{bp+q}{br+s}. \qquad (6)$$

交叉相乘略加变形后得

$$y=\frac{(b-1)(qr-ps)}{b(r-p)+s-q}. \qquad (7)$$

所以，如果（7）中的y是整数，则

$$\frac{p,\ y,\ q}{r,\ y,\ s}$$

是"可约的"。（同定理10中一样，$\frac{p,\ y,\ y,\ q}{r,\ y,\ y,\ s}$也是"可约的"。）（除了2进制外）在一切进位制中都有解。（比如只要

$$\frac{p,\ (y),\ y-p}{r,\ (y),\ y-r}$$

的分子分母都是b进制中的数，它就是解），但我却没有听说有什么普遍性的理论可以使我们不必去考虑相当多的分数。在以10为基时用计算机搜寻，找出了200多个非平凡例（所谓"平凡"，是指$p=q$，$r=s$，或$p=r$，$q=s$或$q=s=0$，或者其倒数已经是解等情况）。3进制中仅只一例（上面已经引用）；4进制中有6例，5进制中有15例。另外，运算过程已表明p与r不必是单位数；所以我们（在10进制中）得到下面两例。

$$\frac{23(9)4}{15(9)6} \quad 和 \quad \frac{18(6)8}{4(6)7}.$$

在10进制中这种例子事实上有2500多个。在3进制和4进制中分别有下面两例：

$$\frac{2,\ 1,\ (2),1}{1,\ 0,\ (2),2} \quad 和 \quad \frac{2,\ 0,\ (3),\ 1}{1,\ 1,\ (3),\ 2},$$

还有许多许多。

现在我们就容易懂得例（F）了。它等价于

$$\frac{143(18)50}{170(18)56},$$

这可以当成100进制中的两个数的商来念，属于（6）的形式，其中p与q是（100进制中的）两位数。我们可以用（7）来核查此例。任何能用电子计算机的人立刻就可以得到更多的例子。

例（E）难懂些，因为我们是约的4个数字中的一个，而且它里面隐藏的代数运算还比较难。这个例子以及可约数字可重复的原理告诉我们还有下例：

$$\frac{3(55)44}{7(55)31},$$

此例可视为又一个100进制例，而且是一个用（7）就可以核实的正确例子。

最后来看（G）。如果用10去乘分子，我们就得到一个10000进制中的两个三位数的商，它们的中间一位数字可以约去。要寻找

$$y=\frac{9999(qr-ps)}{10000(r-p)+s-q}$$

的解，即使用计算机也要耗费不少机时的。在纠正（G）的印刷错误时我假定（倒没有任何特别理由），p和q都正确，然后从900到950来寻找r，从5000到5500中来寻找s。当然我也就找到了（纠正后的）例（G），以及

$$\frac{435(2121)3450}{921(2121)5000},$$

若写成等价形式

$$\frac{425(2121)345}{912(2121)5}$$

样儿就更神秘莫测了。

练习

1．证明定理6。

2．证明定理9。

3．证明（在b进制中）$\frac{x,\ y}{z,\ x}=\frac{y}{z}$的解绝非只是3个。

4．证明若三数组（x，y，z）对b进制是一个解，则对任何别的进位制它就不是解。

5．证明（10进制的）$\frac{pqr}{spq}=\frac{r}{s}$的解可以从（100进制的）$\frac{x,\ y}{z,\ x}=\frac{y}{z}$的解中得出。例：$\frac{847}{484}=\frac{7}{4}$。

6. 请你为$\frac{p, x, q}{r, x, s}=\frac{p, q}{r, s}$ (b进制)创造一个理论。10进制的例上面已经有了，就是 (D)，另外的例子在〔9〕与〔10〕中。

7. 在10进制中再找出 (B) 型的另外那一例。

8. 创造一个理论，把$\frac{695}{139}=\frac{65}{13}=5$等的解释也包括在内。

参考文献

(我感谢让·J.彼得森提供下面文献中的几篇)

1. Anonymous report of a meeting, Scripta Math., 5(1939) 176.

2. R.P.Boas, Anomalous cancellation, Two-Year College Math. J., 3,no. 2(1972) 21—24. (A pre-computer version of the present essay.)

3. R.A.Carman, Mathematical misteaks, Math.Teacher, 64(1971)109—115.(This contains many other curiosities besides cancellations.)

4. A.P.Domoryad, Mathematical Ga-

mes and Pastimes, (Halina Moss, trans.), Macmillan, New York, 1964; Russian edition, Matematicheskie igry irazvlecheniya, GIF-ML, Moscow, 1961.

5. Burton W. Jones, The Theory of Numbers, Rinehart, New York, 1955, 74.

6. W. J. LeVeque, Topics in Number Theory, vol. 2, Addison-Wesley, Reading, Mass., 1956, 201 ff.

7. I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 3rd ed., Wiley, New York, 1972, 100.

8. C. S. Ogilvy and J. T. Anderson, Excursions in Number Theory, Oxford University press, New York, 1966, 86.

9. C. W. Trigg, Solution of problem 434, Math. Mag., 34(1961)367—368.

10. ——, Solution of problem 365, Pi Mu Epsilon J. 6(1977)372—374.

七、几何学的变形观点

G.D.查克里安

1.前言 我们从一张纸上剪一个三角形下来拿到中午的日光下，它的影子可能呈现出一些什么样的形状呢？我们能不能把剪下的三角形摆在某个位置上使其阴影为等边三角形？

我们手上的三角形与它在地上的影像之间的对应关系是名叫**正交射影**的一种重要几何变换的近似的物理实现。著名几何学家丹涅尔·皮多设计并解说的一部很有趣的数学影片〔8〕就是以这种变换为题材的。影片介绍了任意三角形总能安置到某个位置上使其阴影为等边三角形这个事实。

貌似困难的几何问题往往可以通过正交射影变换成更易处理的形式。然后再变回来，更容易解的问题的答案就给出了原问题的答案。（人们也可能会说，这正对应于一条有用的原理，即变形的观点往往对问题带来了更重要的深入的看法。）我们在§2中将会看到在正交射影下，三角形的中线依然投射成三角形的像的中线。由于任意一个三角形总可以经过某个正交射影变成等边三角形(见§2)，

所以我们就知道要证三中线交于一点只需就等边三角形这个特例来证明就够了。诚然，就这个特例来说有点杀鸡用牛刀的味儿，不过耐心的读者在以下各节中会读到更多的难于解决的问题的。

把给定的问题变换成一个容易求解的问题，再解这个容易解的问题，最后变换（反演）回去得到原问题的解，这是一个基本的数学方法．用对数把乘法问题化为加法问题，或者在更高的层次上用拉普拉斯变换把微分方程化为代数方程都是这个方法的经典例子。M.S.克兰姆金和D.J.纽曼的一篇文章〔7〕从数学的各个不同领域里举出了大量例子来阐述这个“变换、求解、反演”方法，很值得好好读一读。豪华德·伊弗斯的著名的几何教科书〔4，130—131页，275—279页〕对于这个方法在几何学中的应用也作了精采的讲解。

我们这儿的目的是对正交射影的某些性质作一番探讨，并把这些性质用来解决多边形内接椭圆方面的某些问题。但是我们并不受这条主线的约束，只要绕道可以到达十分有趣的境界，我们就自由自在地往那里绕一圈。事实上，在本讨论的末尾就对几何不等式作了简短的探讨，稍微偏离了我们的主题。

2.正交射影的性质 试想像一张水平平面π^*以

及π^*上方位于不与π^*垂直的平面π中的图形F。从F的每一点P作π^*的垂线记垂足为P^*，全部这种P^*点所成的集合构成一个图形F^*，叫做F的正交射影。我们可以把F^*看成是π^*上方很远很远处的太阳对F投下的阴影（图36）。

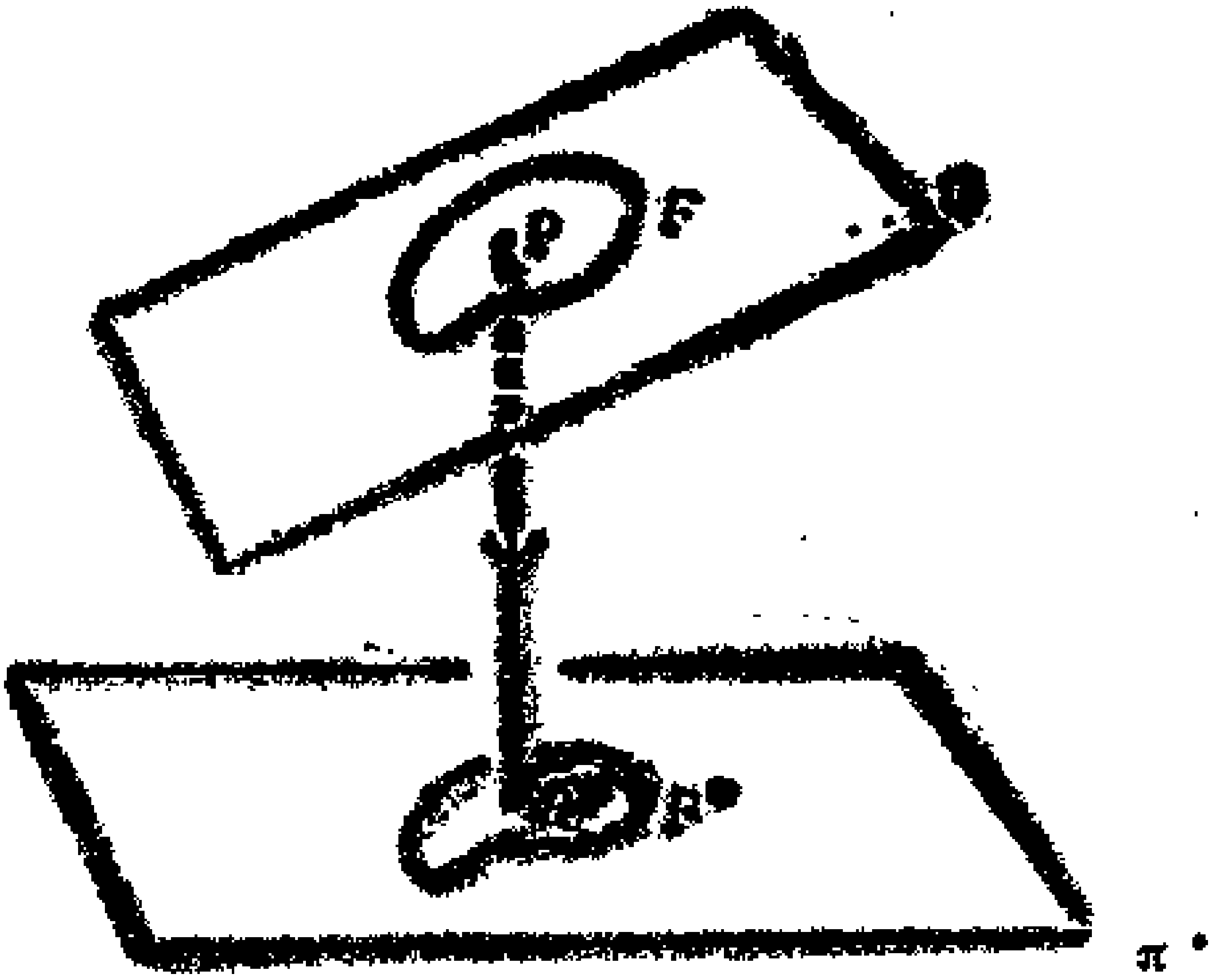


图 36

下面是正交射影的几条基本性质，以后会用到的。

（P1）若l与m是π上的两条平行直线，则其射影l^*与m^*是π^*上的两条平行直线.（"平行线投射成平行线。"）

（P2）同一条直线或两条平行线上的两个线段长度之比，等于其射影长度之比。（"平行线段长度之

比不变。”）

（$P3$）π上两个区域的面积之比等于其射影的面积之比。（“面积比不变。”）

（$P4$）若F是π上的椭圆，则其射影F^*是π^*上的椭圆。而且若给定任意一个椭圆，总能将这椭圆所在平面π的位置加以改变，使这椭圆的射影成为一个圆。

性质（$P1$）是从这样的事实推出的结果，即l^*是π^*与垂直于π^*且含l的平面的交线。因此l^*与m^*是π^*与一对平行平面的交线（图37）。

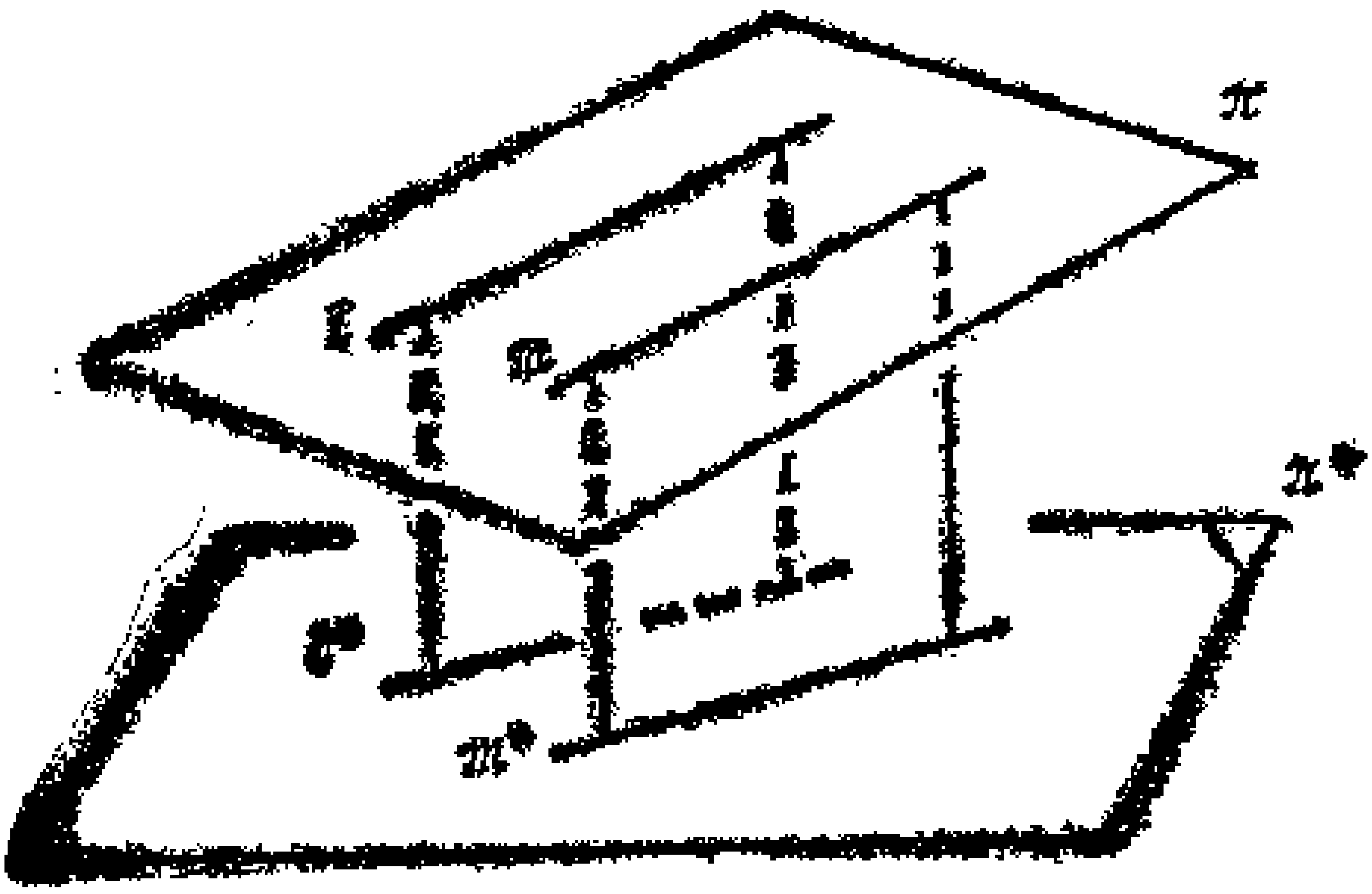


图37

读者定会发现，性质（$P2$）就是相似直角三角形的简易练习。

至于性质（$P3$），我们首先注意，如果π上的两个给定区域是长边垂直于π与π^*交线的窄矩

形，那末性质（$P3$）就可以从（$P2$）很快得到。用这些窄矩形的并集来逼近给定区域就可以得到一般情况的结论。

关于（$P4$）的第二个论断我们可以回想一张平面截一个半径为b的正圆柱，截出一个半短轴为b的椭圆来。关于这个论断的优美证明可以在希尔伯特和柯恩·沃森的几何经典〔5〕的第一章中读到。改变切割平面的倾斜度我们显然可以得到大于b的任意半长轴。所得椭圆的正交射影就是圆柱的圆形底（图38）。

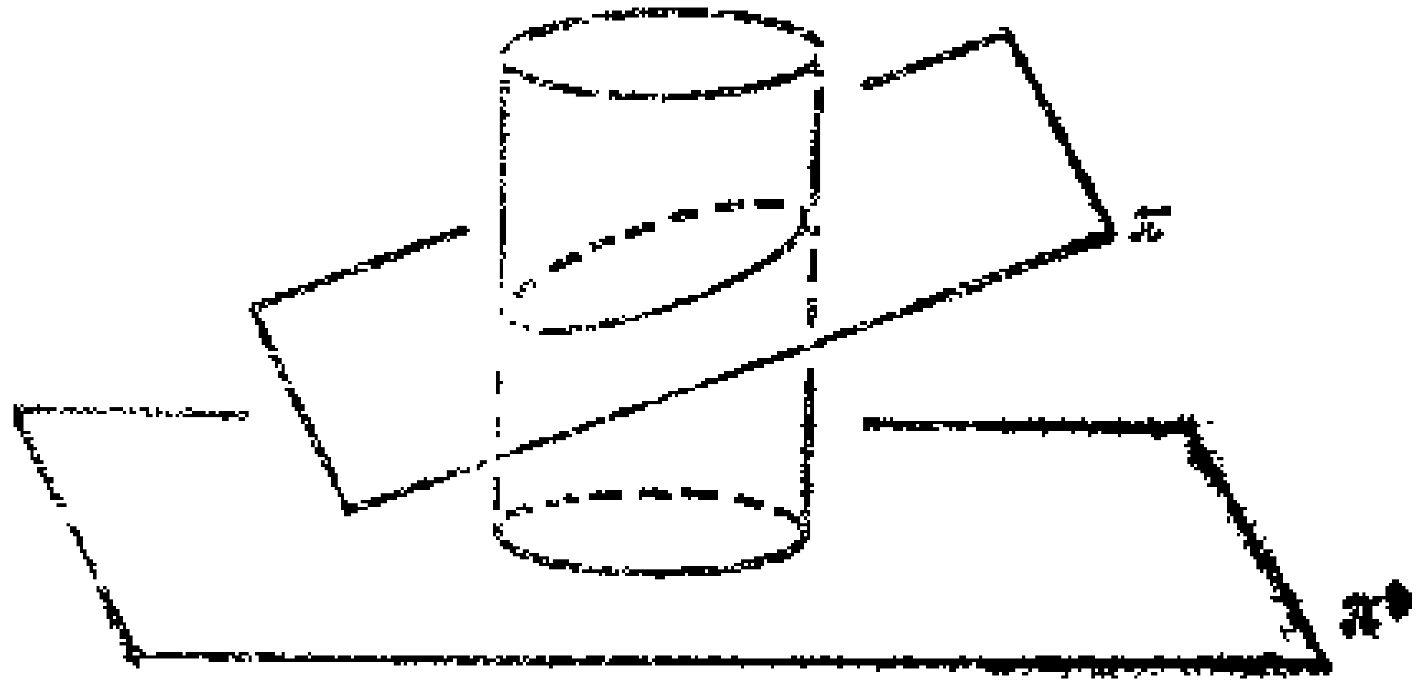


图38

说椭圆永远投射成椭圆，完全等于是说底为椭圆的柱被不平行于其母线的平面所切割出的任意横截面都是椭圆。我们不打算停下来证明这个结论，只是想提醒一下，使用三维空间解析几何〔见练习10〕这门学问就可以证明。关于这个问题，希尔伯特与柯恩·沃森〔5〕书中第一章对于二次曲面

以及它们被平面可能切出的截面都作了非常直观形象的讨论。

性质（$P2$）意味着，若AB是π上的一条线段，其中点为M，则M的正交射影M^*就是AB的正交射影A^*B^*的中点。所以π上的三角形ABC的中线显然都投射成π^*上的三角形$A^*B^*C^*$的中线，这点我们先就说过了。另外，由于三角形的形心是三条中线的交点，所以在正交射影下三角形的形心投射成它的像的形心。

由于线段中点投射成线段的像的中点，所以容易推知，在正交射影下椭圆中心投射成椭圆的像的中心，这一结果以后很有用。

为什么可以调整三角形所在平面的位置使三角形的正交射影是等边三角形，现在正好该给出一个说得通的道理了。我们注意，圆锥曲线的方程含有5个独立常数（我们可以用任意非零常数乘方程而不会改变方程所表达的曲线），我们有一个简略规则：圆锥曲线由5个条件唯一地确定。可以说在画圆锥曲线时我们有5个自由度。如果已给一个三角形ABC，按此规则我们知道有一条唯一的圆锥曲线切BC边于其中点P，切CA边于其中点Q，切AB边于某点（要求圆锥曲线与三条边相切，用去三个自由度；要它过两点P、Q又消耗两个自由度，总共用去5个自由度）。虽然我们不打算在这儿进行

证明，但是可以证实，满足这些条件的圆锥曲线E是的确存在的，而且它就是一个椭圆。我们现在用正交射影要证明的是，若R是E与边AB相切的切点，则R事实上必定是AB的中点。

性质（$P4$）告诉我们，三角形ABC所在的平面π可以安排来使椭圆E的正交射影E*成为一个圆（图39）。由于在正交射影下相切性不变，ABC

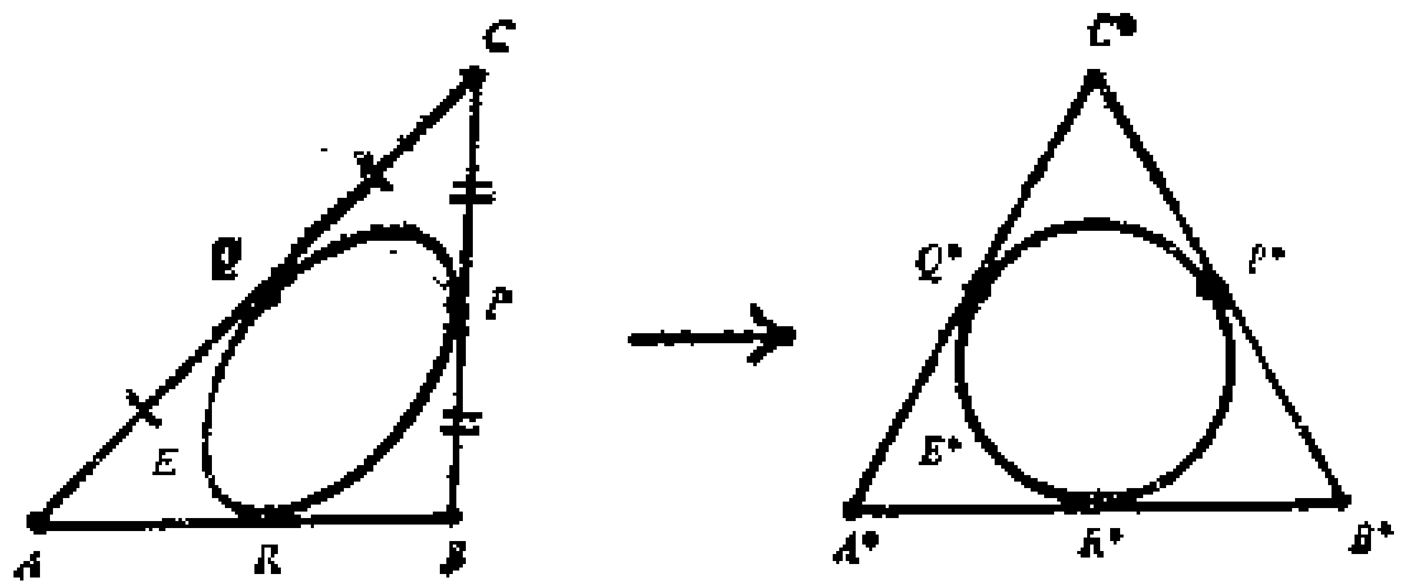


图39

的射影是外切于圆E^*的三角形$A^*B^*C^*$，边B^*C^*切于其中点P^*，边C^*A^*切于其中点Q^*，A^*B^*切于R^*。由于圆外一点到圆的两条切线段等长，所以我们有

$A^*Q^*=A^*R^*$，$B^*R^*=B^*P^*$，$C^*P^*=C^*Q^*$。但是$B^*P^*=C^*P^*$，$C^*Q^*=A^*Q^*$，所以上面几个等式意味着R^*是A^*B^*边的中点，$A^*B^*C^*$是等边三角形。

我们一下子就同时发现了两件事：即在一个已给的三角形中有唯一的一个椭圆存在，它与三边相

切于各中点；把这个椭圆投射成圆的正交射影把三角形投射成一个等边三角形. 在内切于三角形的诸椭圆中这个中点椭圆具有特殊的地位. 由于等边三角形$A^*B^*C^*$的形心就是其内切圆的圆心，所以ABC的形心也就是其中点椭圆的中心. 以后我们还会看到，中点椭圆的特殊之处就是它是诸内切椭圆中面积最大的一个. 读者在皮多教授的令人思想活跃的文章〔9〕中，将会读到中点椭圆的一个有趣的应用例子，其思路和我们下面的讨论是一致的. 至于中点椭圆的实样，可以在获奖数学影片"中心对称"〔2〕中看到.

3.四边形的内切椭圆 给定的凸四边形的内切椭圆中心的轨迹的确定问题，牛顿解决得很优美. 这个问题在多列的书〔3，p. 217〕中讲到了；若用正交射影来解也是很令人惊叹的.

我们来看已给凸四边形$ABCD$的内切椭圆族，即$ABCD$内的与其四边相切的那些椭圆. 规定椭圆要与四条边相切，我们就用掉了4个自由度. 根据我们的简易规则，即圆锥曲线由5个约束条件唯一地确定，可以预料一定有一个"一维"的椭圆族内切于$ABCD$，与剩下的一个自由度相对应. 我们可以预期内切于$ABCD$的椭圆中心轨迹是某种曲线. 我们事实上将证明的是，内切于四边形ABCD的椭圆

中心全都在$ABCD$的对角线中点连结成的直线段上（图40）。

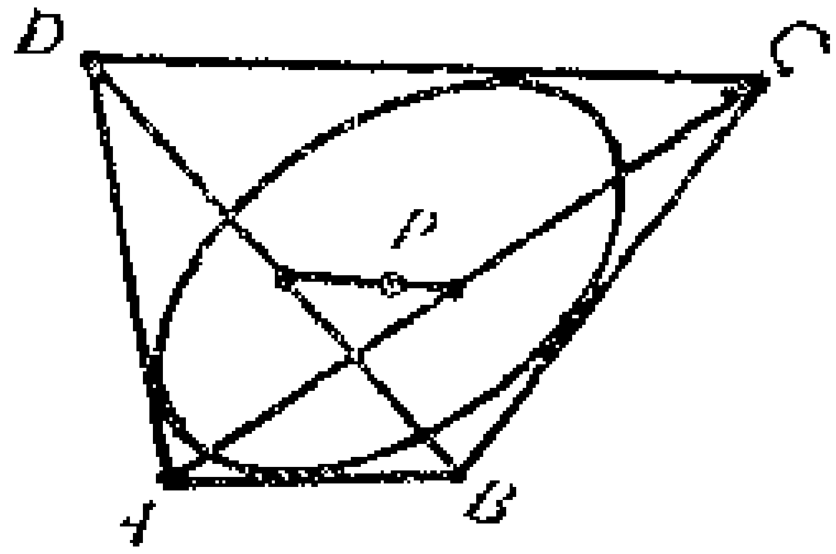


图40

令P为内切于$ABCD$的一个椭圆的圆心．存在一个正交射影，使椭圆投射成圆，ABCD投射成外切此圆的四边形$A^*B^*C^*D^*$．椭圆中心P投射成圆心P^*（图41）

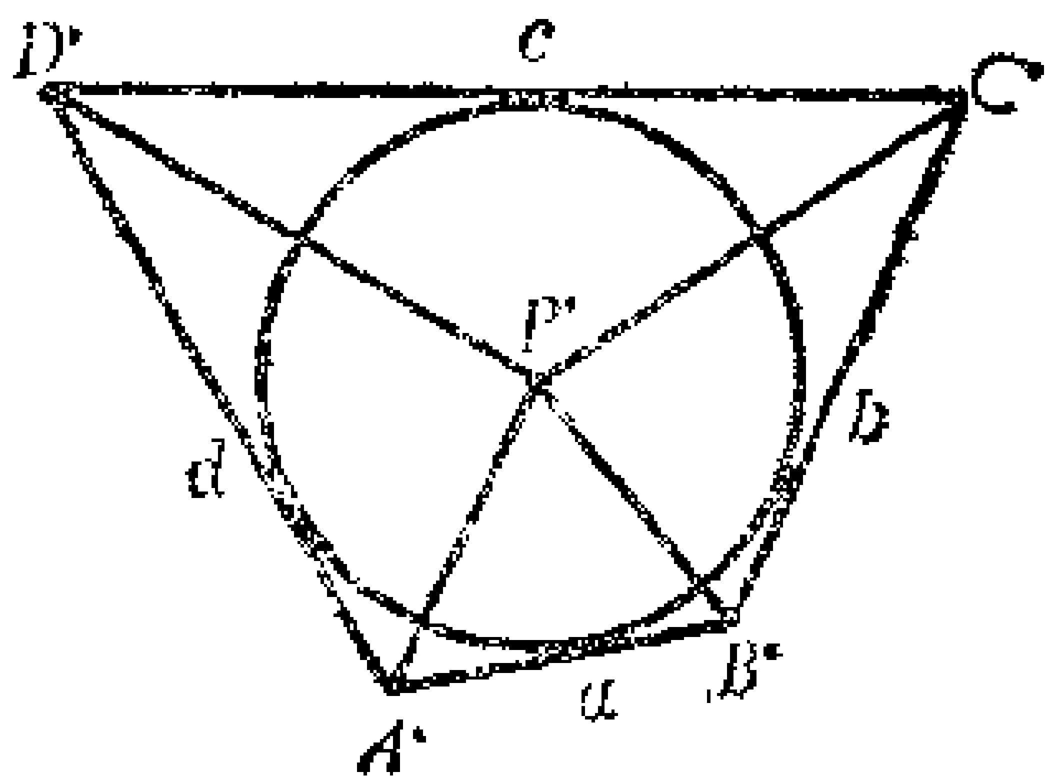


图41

令r为圆的半径，用a，b，c，d分别表示A^*B^*，

B^*C^*, C^*D^*，D^*A^*各边的边长。注意

$$a+c=b+d \qquad (1)$$

外切于圆的四边形的这个性质即圆外一点到圆的两个切线段长度相等，也可以从用来讨论中点椭圆的同一事实推出。方程（1）通乘以$\frac{r}{2}$得

$$\text{面积}A^*P^*B^*+\text{面积}C^*P^*D^*$$
$$=\text{面积}B^*P^*C^*+\text{面积}D^*P^*A^*. \qquad (2)$$

面四个三角形面积之和等于面积$A^*B^*C^*D^*$，故

$$\text{面积}A^*P^*B^*+\text{面积}C^*P^*D^*$$
$$=\frac{1}{2}\cdot\text{面积}A^*B^*C^*D^* \qquad (3)$$

即两个相对的三角形占了四边形面积的一半。运用性质（$P3$）我们可以说，若P是四边形$ABCD$的内切椭圆的中心，则

$$\text{面积}APB+\text{面积}CPD=\frac{1}{2}\text{面积}ABCD \qquad (4)$$

正交射影带来了关键方程（4），剩下要证的是，满足这个方程的那些点P都在一条直线上。为方便起见在$ABCD$所在的平面上引入（x，y）坐标系。令$kx+ly+e=0$为过A，B的直线的方程，$mx+ny+f=0$为过C，D的直线的方程；d_1为从P到AB边的距离，d_2为从P到CD边的距离。如果P点的坐标是（x，y），那么用点到直线的距离公

式得

$$d_1=\pm\frac{kx+ly+e}{\sqrt{k^2+l^2}},\quad d_2=\pm\frac{mx+ny+f}{\sqrt{m^2+n^2}} \tag{5}$$

我们有面积$(APB)=\frac{1}{2}\alpha d_1$，面积$(CPD)=\frac{1}{2}\beta d_2$，式中$\alpha$为$AB$边长，$\beta$为$CD$边长。令$\gamma=(ABCD)$的面积，则方程（4）成为

$$\alpha d_1+\beta d_2=\gamma. \tag{6}$$

用（5）的d_1，d_2代入上式，并记住α，β，γ是常数，我们便知道（6）是P的坐标x，y的线性方程，（P变化时它的$\pm$号不变，因为P不过AB边或CD边）。所以P始终在与四边形$ABCD$相关连的某条固定直线上。证毕。

若P是$ABCD$的任一条对角线的中点，则它满足（4）；验证这一点读者不会有任何困难。于是（6）代表的直线上的确含有$ABCD$两条对角线的中点。证明任意内切椭圆的中心都位于两条对角线中点之间的任务，作为一道难题留给读者作练习（见练习4）。

把这两条对角线看成是内切于四边形的“退化”椭圆是很有启发意义的。两条对角线的中点是两个退化椭圆的中心。

最后，依然用退化的说法，聪明的读者可能早已看出两条对角线中点重合时也出现退化的情况。

这时四边形是一个平行四边形，所有的内切椭圆的中心都在同一点，即平行四边形的中心。

4．三角形的内切椭圆 确定已给三角形内切椭圆中心的轨迹，也可以用正交射影法来解决。乍一看这个问题不会是很激动人心的，但是我们将会发现这个问题的解决还是动用了许多十分有趣的几何思想。

我们来看已给三角形ABC的内切椭圆族。要使椭圆与三角形三边相切就用掉三个自由度。确定圆锥曲线时的简略规则使我们产生了希望，希望有一个内切于ABC的"二维"椭圆族存在，它对应于剩下的两个自由度。于是我们期望内切于ABC的椭圆中心的轨迹是某种区域。如果证明了：**内切于三角形ABC的椭圆中心的轨迹是ABC三边中点所确定的三角形的内部区域**，也就实现了我们的希望。事实上我们将证明，以ABC三边中点作为顶点的三角形的每个内点都是内切于ABC的唯一一个椭圆的中心。

令D、E、F为三角形ABC三边的中点，如图42所示。要证明ABC的任意（非退化的）内切椭圆的中心P都在三角形DEF之内，就是件很平常的事了，事实上P到AB边的距离等于AB与椭圆的平行于AB的另一切线之间距离d的一半；而d是小于C到AB的距离的。所以P到AB的距离小于C到AB

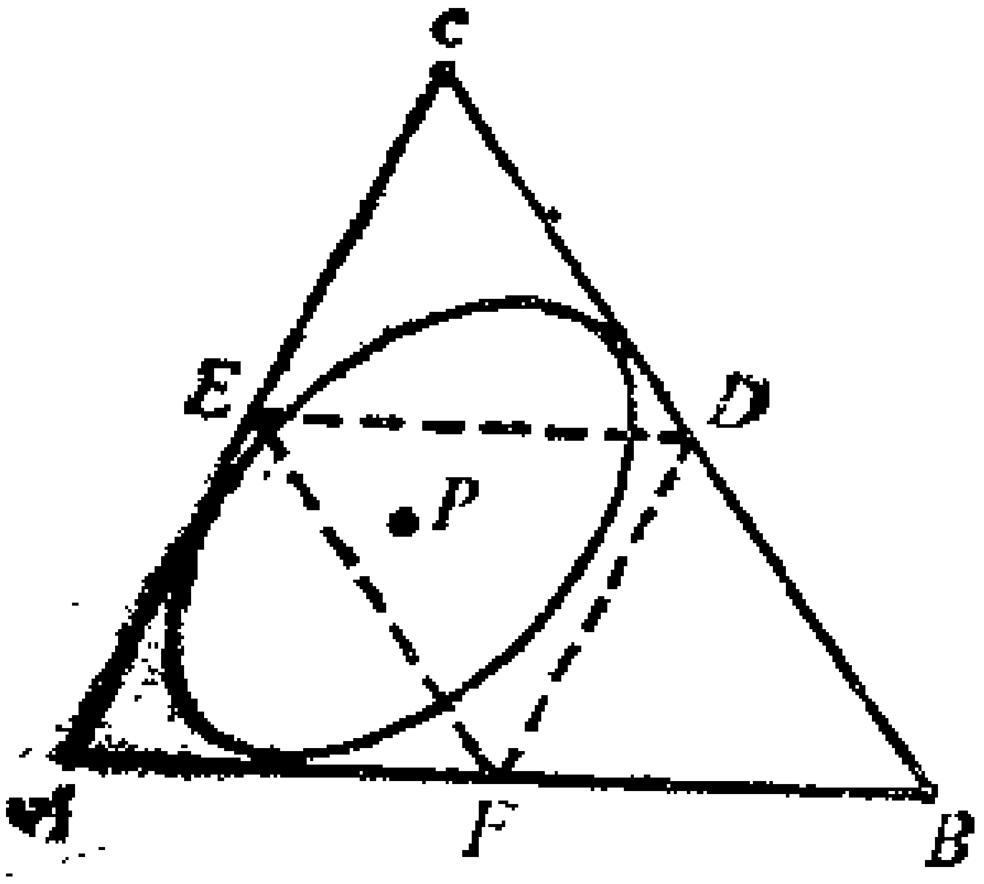


图42

的距离的一半。故P与C在直线DE异侧。同理可知P与A在EF的异侧，P与B在FD的异侧。因而P在DEF内。

（上面讲的道理完全实用于内切于ABC的任何中心对称的凸曲线。如果K是一条固定的中心对称凸曲线，我们可以定义K曲线类为K经过正交射影与尺度改变得到的那些凸曲线。所以若K为圆，则K曲线类就是椭圆。切望开辟新领域的读者可能想知道关于内切椭圆的讨论推广到对内切K曲线的情况。）

我们力图证明DEF的任意一个内点都是内切于ABC的某个椭圆的中心，这番努力会使我们兴趣大增。

我们知道只讨论ABC为等边三角形的特别情况

就够了，这样一来问题可以大大地简化。事实上若ABC为一个任意三角形，P为其中点三角形DEF的内点，则ABC可以通过正交射影投射成等边三角形$A^*B^*C^*$，P可以投射成$A^*B^*C^*$的中点三角形$D^*E^*F^*$的一个内点P^*。若此时能找出一个中心在P^*而又内切于$A^*B^*C^*$的椭圆，那么它正是我们所求的中心在P并内切于ABC的椭圆的射影。

因此我们只注意等边三角形ABC，研究其中点三角形DEF的内点P。令α为三角形BPC的面积，β为CPA的面积，γ为APB的面积，则P的位置完全由三数组(α, β, γ)确定。如果假设三角形ABC面积为1，则α，β，γ就是点P关于ABC的所谓重心坐标。为了能确信ABC内两个不同点不可能有相同的重心坐标，我们来看一下有一个坐标是常数的点的轨迹，这轨迹是与三角形的一条边相平行的直线（因为底相同的等面积三角形具有相同的高；）而且两条不同的直线至多交于一点。

暂且假设我们可以找到一个三角形$A'B'C'$，它的内切圆圆心P'满足下式

$$\text{面积}(B'P'C'):\text{面积}(C'P'A'):\text{面积}(A'P'B')=\alpha:\beta:\gamma. \qquad (7)$$

那么用适当的正交射影就可以把$A'B'C'$投射成等边三角形ABC，把其内切圆投射成ABC的内切椭圆，把P'投射成椭圆的中心P。根据性质（$P3$）有

$$面积(BPC):面积(CPA):面积(APB) = \alpha:\beta:\gamma. \qquad (8)$$

换言之，我们得到等边三角形ABC和中心在点P的内切椭圆，点P的重心坐标与已知坐标α，β，γ成比例，把比例尺度简单一变就得到我们问题的解。

于是对我们的问题的解决便归结为作出下面的的证明。令ABC为已知等边三角形，P为其中点三角形内的一点。令$\alpha=(BPC)$的面积，$\beta=(C\ \ A$的面积，$\gamma=(APB)$的面积。那么存在一个三角形$A'B'C'$，其内切圆心P'满足关系式(7)。

证明这一点的关键在于，α，β，γ不是任意的正数。我们必须用到P在ABC的中点三角形内这一事实。而对于这样的点P，若α，β，γ如上定义，则有

$$\alpha+\beta>\gamma,\ \beta+\gamma>\alpha,\ \gamma+\alpha>\beta, \qquad (9)$$

这是惹人喜欢的精巧问题，它的证明姑且留给读者作为一个简单的练习吧（见练习5）。关系式(9)正好是使边长为$B'C'=\alpha$，$C'A'=\beta$，$A'B'=\gamma$的三角形$A'B'C'$存在所需的条件。若P'是这个三角形内切圆的圆心，r是半径，则

$$面积(B'P'C')=(1/2)r\alpha,$$

$$面积(C'P'A')=(1/2)r\beta,$$

$$面积(A'P'B')=(1/2)r\gamma, \qquad (10)$$

故关系式(7)成立，问题解决。

至此我们已经证明了三角形ABC的每个内切椭圆的中心在其中点三角形内部，反之在中点三角形内部的每个点是ABC的一个内切椭圆的中心。但是中心在点P并内切于ABC的椭圆最多只能有一个。（要明白这一点的最好办法也许是注意一下两个同心椭圆的公切线，如果如图43所示那样相交必定构成一个平行四边形）。于是与中点三角形的每个内点P有唯一的一个以P为中心并内切于三角形ABC的椭圆与之一一对应。所以关于ABC的内切椭圆的问题常常可以用中点三角形内的点来表达。下一节推导的面积公式就充分说明了这一点。

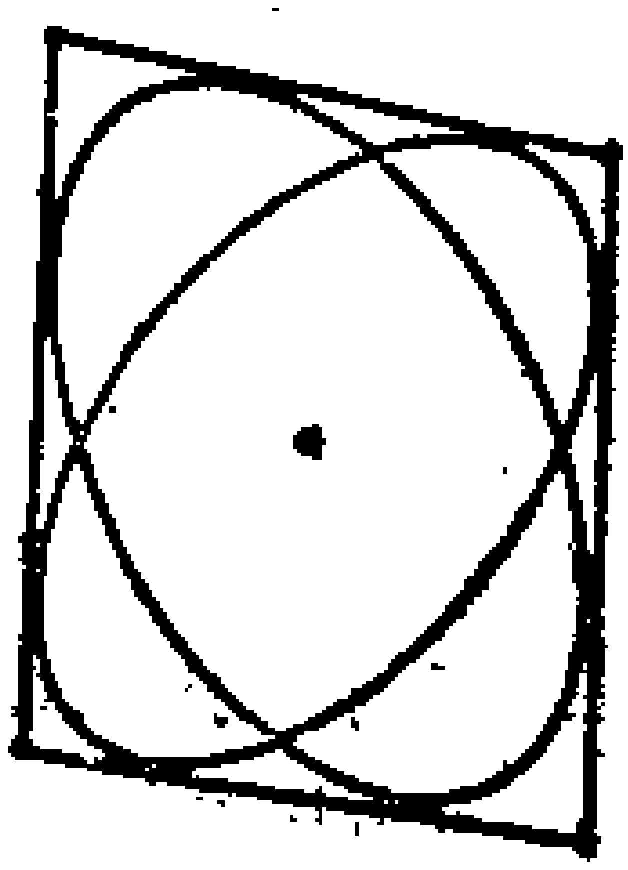


图43

5.内切椭圆的面积 若ABC是任意一个面积为1的三角形，则ABC内的每一点P都有重心坐标(α, β, γ)可以与之相联系，这坐标由下式给出：

$$\alpha = (BPC) \text{ 的面积,}$$

$$\beta = (CPA) \text{ 的面积}, \tag{11}$$
$$\gamma = (APB) \text{ 的面积}.$$

和前节一样我们看出ABC中的每一点P都唯一地由其重心坐标确定。而且若P是ABC的中点三角形内的点，则有唯一的一个椭圆存在，其中点在P且内切于ABC。所以完全有理由相信这个椭圆的面积可以用它的中心P的重心坐标来表达。

事实上内切于具有单位面积的三角形ABC的椭圆E，当其中心P的重心坐标为（α，β，γ）时，面积由下式给出：

$$(E)\text{的面积} = \pi[(1-2\alpha)(1-2\beta)(1-2\gamma)]^{\frac{1}{2}} \tag{12}$$

证明这个公式主要靠正交射影。

设P为ABC的中点三角形内的一点，其重心坐标(α，β，γ)由（11）式给出。用一个适当的正交射影可以把中心在P且内切于ABC的椭圆E投影成半径为r、圆心为P^*且内切于$A^*B^*C^*$的圆E^*

由性质（$P3$）可知

$$\text{面积}(B^*P^*C^*) : \text{面积}(C^*P^*A^*) : \text{面积}(A^*P^*B^*) = \alpha : \beta : \gamma \tag{13}$$

但面积 $(B^*P^*C^*) = (\frac{1}{2}r)(B^*C^*)$，

面积 $(C^*P^*A^*) = (\frac{1}{2}r)(C^*A^*)$，

面积 $(A^*P^*B^*) = (\frac{1}{2}r)(A^*B^*)$，故由

(13) 可知存在一个常数k，使得

$B^*C^*=k\alpha$，$C^*A^*=k\beta$，$A^*B^*=k\gamma$，　(14)

如图44所示。

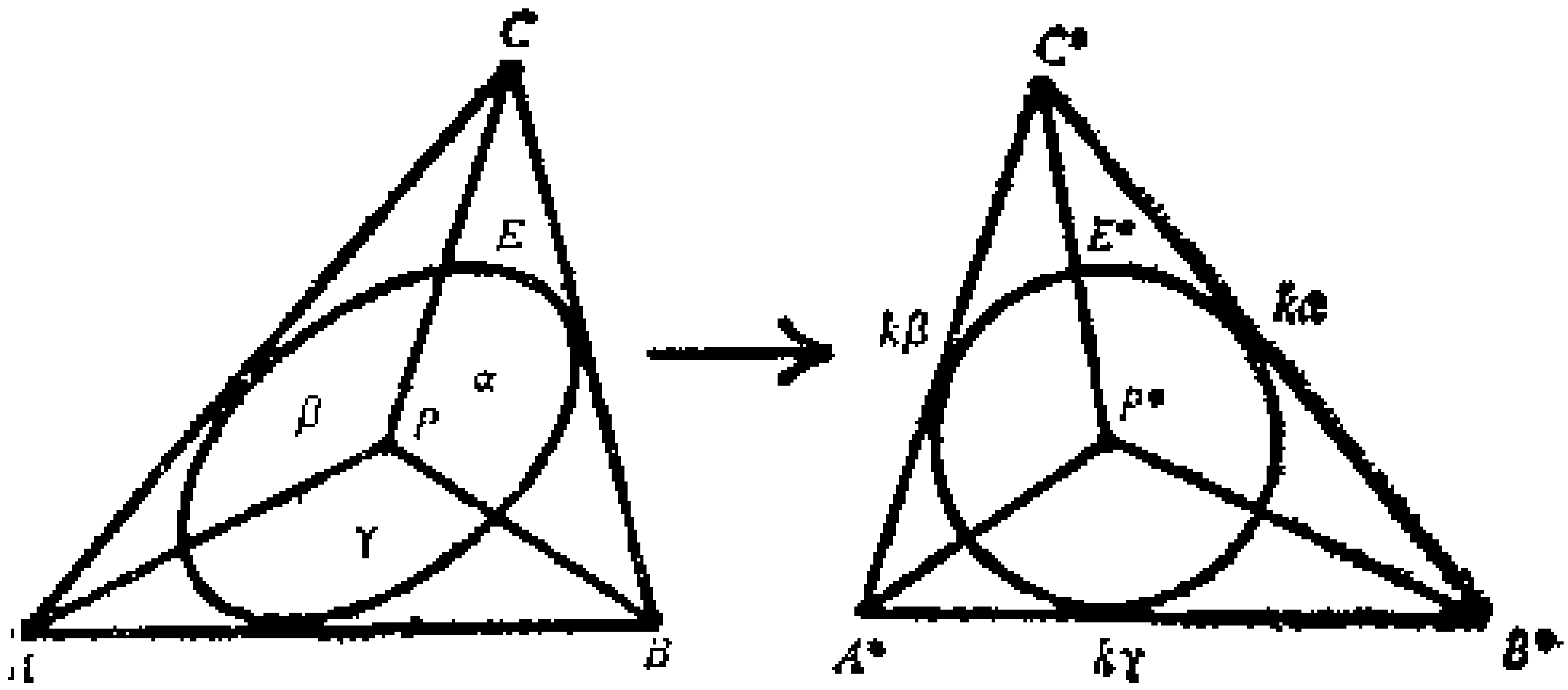


图44

三角形$A^*B^*C^*$的周长之半$s=\left(\frac{1}{2}\right)(k\alpha+k\beta+k\gamma)=\left(\frac{1}{2}\right)k$（因为$ABC$具有单位面积，故$\alpha+\beta+\gamma=1$）。运用由三角形的边长表达三角形面积的古老的海伦公式，得

$$\text{面积}(A^*B^*C^*)=\left[\left(\frac{1}{2}k\right)\left(\frac{1}{2}k-k\alpha\right)\cdot\left(\frac{1}{2}k-k\beta\right)\left(\frac{1}{2}k-k\gamma\right)\right]^{\frac{1}{2}}$$

$$=\left(\frac{1}{4}k^2\right)\left[(1-2\alpha)(1-2\beta)(1-2\gamma)\right]^{\frac{1}{2}}\quad(15)$$

$A^*B^*C^*$的面积当然也由下式给出：

$$\text{面积}(A^*B^*C^*)=\left(\frac{1}{2}r\right)(k\alpha+k\beta+k\gamma)$$

$$=\frac{1}{2}rk.$$

(16)

由(15)与(16)消去k得到

$$\frac{\pi r^2}{面积(A^*B^*C^*)}$$

$$=\pi[(1-2\alpha)(1-2\beta)(1-2\gamma)]^{\frac{1}{2}} \qquad (17)$$

由性质（$P3$）和面积（ABC）＝1推出

$$面积(E)=\frac{面积(E)}{面积(ABC)}=\frac{面积(E^*)}{面积(A^*B^*C^*)}$$

$$=\frac{\pi r^2}{面积(A^*B^*C^*)} \qquad (18)$$

最后合并（17）和（18）得到要证的公式（12）。

注意，（12）中的因子$1-2\alpha$，$1-2\beta$，$1-2\gamma$都是正的，因为P在中点三角形之内。

6．中点椭圆的面积 由我们的内切椭圆面积公式还能推出一个有趣的结果。我们可以用这个公式证明：**面积为1的已知三角形的全部内切椭圆之中，§2所定义的中点椭圆具有最大的面积。**

证明中要用到数学中最重要的不等式之一，即算术平均与几何平均之间的著名不等式。这个不等式说，n个正实数的几何平均值绝不会大于它们的算术平均值，而且只有当所有这些数都相等时几何平均才等于算术平均。在三个正实数a，b，c的情况这

个不等式就是

$$(abc)^{\frac{1}{3}} \leqslant \frac{a+b+c}{3}, \tag{19}$$

等号仅在$a=b=c$时成立。一般情况下的这个不等式多年来有了许许多多不同的证明(参看〔1〕,〔6〕).

(19)这个特殊情况有一个巧妙的证明，它的优点是简洁，但也有个缺点，就是显得象是从代数帽子里变出一个数学兔子来。我们且来看一下，如果x，y，$z>0$，那么恒等式

$$x^3+y^3+z^3-3xyz=\frac{1}{2}(x+y+z)[(x-y)^2+(y-z)^2+(z-x)^2] \tag{20}$$

的右边除了$x=y=z$外显然为正。因而左边是非负的，而且只在$x=y=z$时才等于0。当$a=x^3$，$b=y^3$，$c=z^3$时经过简单运算便得到(19)。

我们把(19)运用到公式(12)，令$a=1-2\alpha$，$b=1-2\beta$，$c=1-2\gamma$，由于$a+b+c=3-2(\alpha+\beta+\gamma)=1$，所以关系式(19)的形式为

$$[(1-2\alpha)(1-2\beta)(1-2\gamma)]^{\frac{1}{3}} \leqslant \frac{1}{3}, \tag{21}$$

等号仅当$\alpha=\beta=\gamma=1/3$时才成立。应用(12)我们看到若E是内切于单位面积三角形的椭圆，E的中心有重心坐标(α，β，γ)，则

$$(E)\text{的面积} \leqslant \frac{\pi}{3\sqrt{3}}, \tag{22}$$

只有$\alpha=\beta=\gamma=1/3$时等号才成立。形心是三角形重

心坐标相等的唯一一个点，中心在该点的唯一内切椭圆是中点椭圆．所以具有单位面积的三角形的中点椭圆的面积是$\pi/3\sqrt{3}$，而其它椭圆的面积都小一些．证毕．

7．一个等周定理　中点椭圆和内切圆是内切于已知三角形的特别有趣的两个椭圆，所以把二者加以比较是很自然的事．内切圆面积不大于中点椭圆的面积这个性质把我们引向一个重要定理，三角形的等周定理．

设Ⅰ是具有单位面积的三角形ABC的内切圆．若Ⅰ的半径是r，则（Ⅰ）的面积$=\pi r^2$，（ABC）的面积$=\left(\frac{r}{2}\right)L$，此处$L$是$ABC$的周长．由此二式消去$L$得

$$\text{面积}(\mathrm{I})=\frac{4\pi}{L^2}[\text{面积}(ABC)]^2=\frac{4\pi}{L^2}. \quad (23)$$

由（22）我们有面积（Ⅰ）$\leqslant\frac{\pi}{3\sqrt{3}}$结合(23)得到

$$L^2\geqslant 12\sqrt{3}. \quad (24)$$

(24)中的等式仅当Ⅰ也是ABC的中点椭圆时才成立，此时ABC必定是一个等边三角形．心中记住比数L^2/T在变换尺度时是不变量，此处L是周长，T是面积，我们就可从(24)推出不等式

$$L^2\geqslant 12\sqrt{3}\,T, \quad (25)$$

此式对周长为L，面积为T的任意三角形都成立。另外，(25)中的等号只当三角形是等边三角形时才成立。

略加思索就可以证明，不等式(25)连同等号成立的条件，有一个等价的论断，即周长给定的全部三角形中，等边三角形面积最大。这个论断是对于三角形而言的等周定理。

刚才讲的是n边形的等周定理的特例，一般n边形的情况是：周长给定的全部n边形中，等边n边形面积最大。

这方面最有名的定理是一般等周定理：具有给定周长的所有简单封闭平面曲线中，圆的面积最大。

我们的漫游就在几何不等式这个主题的门口停下来了，在这里我们最后碰见的定理是这门学科中最令人肃然起敬又最有吸引力的。想游览到更深处去的读者在N.D.卡札林诺夫的专论〔6〕中将会找到这方面的非常出色的初等的介绍。

练习

1.假定把三角形三边上各一点互相连结而使其分割成四个小三角形。试证内三角形的面积不可能小于任何一个外三角形的面积（图46）。〔提示：

用正交射影把内三角形投射成一个等边三角形。〕

2.用正交射影来证明椭圆的一组平行弦上的中点是共线的。

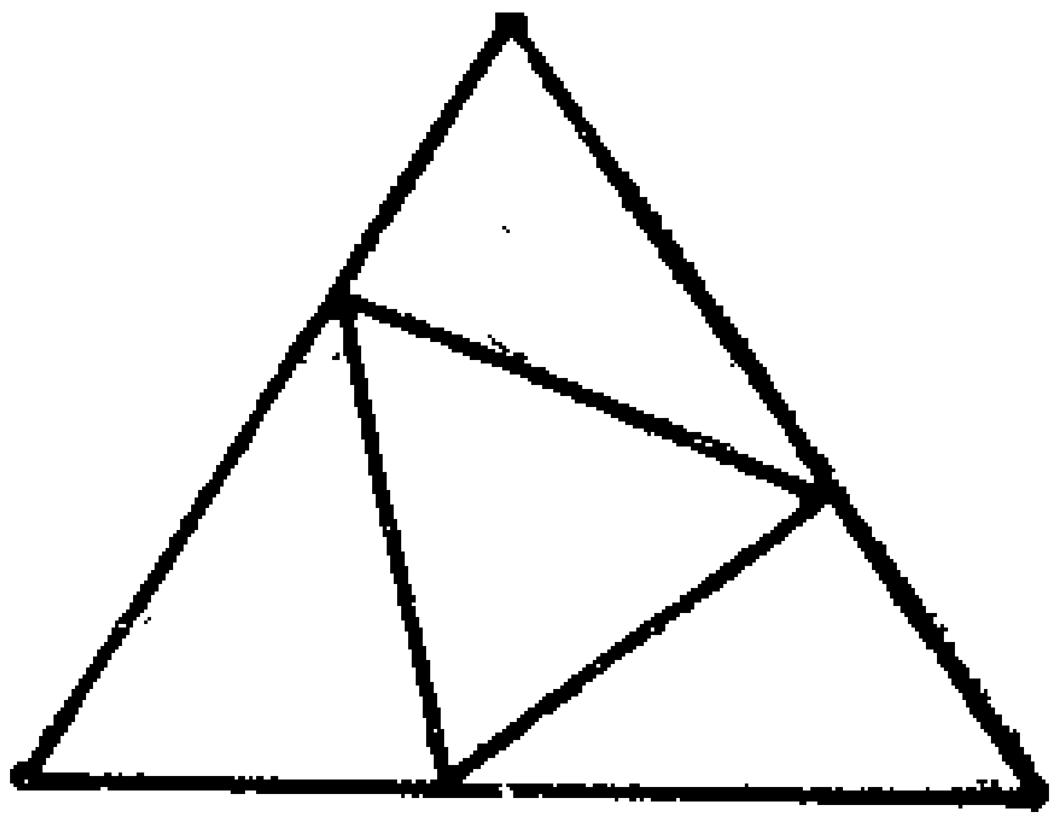


图45

3.试证：内接于椭圆的三角形有最大面积的充要条件是其形心与椭圆中心重合。

4.试证凸四边形内切椭圆的中心永远在对角线的两个中点之间。〔提示：用正交射影把问题化为凸四边形$ABCD$的内切圆的问题。令P为圆心。只消证明P和AC的中点在BD的同一侧，P和BD的中点在AC的同一侧。令l为过P点且垂直于AC的直线。设AC的中点M不在l上，证明l把M从B和D分隔开了。〕

5.若(α, β, γ)是属于中点三角形的一点的重心坐标，试证

$$\alpha+\beta>\gamma,\ \beta+\gamma>\alpha,\ \gamma+\alpha>\beta.$$

6. 已知一个三角形ABC，其面积不一定等于1；以及中点三角形内一点P，并令α，β，γ由（11）式定义。令E为中心在P并内切于ABC的椭圆。令$\sigma=(\frac{1}{2})(\alpha+\beta+\gamma)=\frac{1}{2}$面积（$ABC$），请将公式（12）推广成

$$\text{面积}(E)=\frac{4\pi}{\text{面积}(ABC)}[\sigma(\sigma-\alpha)(\sigma-\beta)\cdot(\sigma-\gamma)]^{\frac{1}{2}}$$

7. 设有一椭圆内切于三角形ABC，切BC于P，切CA于Q，切AB于R。试证AP，BQ，CR三线共点［提示：根据彻伐(Ceva)定理，这三线共点的充要条件是$(AR)(BP)(CQ)=(RB)(PC)(QA)$。］

8. 三角形ABC内接于椭圆E。令α，β，γ为E的过中心且分别平行于BC，AC，AB的三条弦的长度。令R为ABC的外接圆的半径。试证

$$R=\frac{\pi}{8}\ \frac{\alpha\beta\gamma}{\text{面积}(E)}.$$

［提示：若三角形三边边长为a，b，c，面积为T，外接圆半径为R，则$T=abc/4R$。考虑把E投射成圆的正交射影，把上述结论用到ABC的像上去。］

9. 设X是平面曲线C上的一个点，Y，Z为C上的邻近两点。若沿着C令$Y\to X$，$Z\to X$，并对C附加适当的假设条件，则XYZ的外接圆趋于一个过X

并切于C的圆。这个极限圆从某个意义上来说是在X点用圆对C的最佳逼近，这个圆叫X点的密切圆。X点的密切圆的半径叫做C在X的曲率半径。

令X为椭圆E上的一点，d为过E的中心且平行于在X点的切线的弦的长度。若R为E在X处的曲率半径。试用前一个练习题证明

$$R=\frac{\pi}{8}\ \frac{d^3}{\text{面积}(E)}.$$

可以完全等价地说，若r是E的与X点处的切线平行的半径，E的二轴为a与b，则

$$R=\frac{r^3}{ab}.$$

10. 设在(x, y, z)空间的任意位置椭圆柱面的方程是x，y，z的二次多项式。由此证明：椭圆柱面被不平行于其母线的平面截出的截口是一个椭圆。

参考文献

1. E. F. Beckenbach and R. Bellman, An Introduction to Inequalities, Math. Assoc. of America, New Mathematical Library, No. 3.

2. Central Similarities, University of Minnesota College Geometry Project, distr-

ibuted by International Film Bureau, Inc.

3. H. Dorrie, 100 Great Problems of Elementary Mathematics,Dover, New York, 1965.

4. Howard Eves, A Survey of Geometry, rev.ed.,Allyn and Bacon, Boston, 1972.

5. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.

6. N. D. Kazarinoff, Geometric Inequalities, Math. Assoc. of America, New Mathematical Library, No. 4.

7. M.S. Klamkin and D. J. Newman, The philosophy and applications of transform theory, SIAM Review, 3 (1961) 10-36.

8. Orthogonal Projection, University of Minnesota College Geometry Project, distributed by International Film Bureau, Inc.

9. Daniel Pedoe, Thinking geometrically, Amer. Math. Monthly, 77 (1970) 711-721.

八、收敛、发散与计算机

R·P·波亚士

我认为，我们中间那些必须给学生讲授无穷级数的人过分喜欢使用正项级数收敛性判别法。也许错误是在教科书身上。可以证明（这是有益的练习）比值判别法可以更进一步精细化而成为拉阿北判别法，这个判别法说，如果 $\varepsilon(n)\longrightarrow 0$，$h>1$ 时 $1-(a_{n+1}/a_n)>\dfrac{h+\varepsilon(n)}{n}$，则 $\sum a_n$ 收敛（真的，任何能独自证明这一点的学生无疑地都懂得“收敛”的定义），但是这个判别法只在很特殊的场合下才有用。事实上如果一个级数收敛得快，几乎任何判别法都起作用；如果级数收敛得慢，我们不大可能奈它何，即便用上高速计算机来帮忙，我们将会看到想把一项接一项地加起来求和也奈何不得。错把“判别法”过分强调的原因，我想是由于对收敛（或发散）究竟是怎么回事缺乏了解。

问题部分是由于我们的思想方法似乎主要是在幂级数基础上形成的。幂级数很典型，如果收敛的话那就收敛得很快，如果不收敛那就发散得很快。

就以$\sum_0^\infty x^n$，$x=0.99$为例吧：这x很接近收敛区间的端点。我们很幸运，对部分和$s_n=\sum_{n=0}^{N} x^n$我们有一个简洁的公式，即

$$s_n=\frac{1-x^{N+1}}{1-x},$$

所以，当$|x|<1$时，和$s=1/(1-x)$；因而就余项来说有$s-s_n=x^{N+1}/(1-x)$。所以我们很容易计算出确保$s-s_n<0.005$（大致是二个十进位的精度）或$s-s_n$小于任何一个我们想考虑的其它数时所需的项数$N+1$，例如取$x=0.99$时当（且仅当）$(0.99)^{N+1}/(0.01)<0.005$时，

即

$(0.99)^{N+1}<5\times10^{-5}$,

$(1/0.99)^{N+1}>20000$,

$$N+1>\frac{\log20000}{\log(1/0.99)}=985.4,$$

时我们才有$s-s_N<0.005$，所以需要有986项。同样的x取误差小于5×10^{-100}，就需要23209项，加这么多项用手算太烦人，但用计算机就容易。

但是其它类型的级数就不像上面的那么好对付了。特别富有启发性的例子是$\sum_1^\infty 1/n^2$，这个级数

作例是很方便的，因为已经知道它的和是$\pi^2/6$，所以只要我们想求很多很多位和，我们就能够把和求到那么多位（因为π的10^5位小数都公布了，至少10^6位已经知道了）。但是如果想凭手算，一项项地求级数和达到很高的准确度，我们会劳很大的神。我们不得不求200项之和才会使误差小于0.005（这要求不算不合理吧），但每增加一位小数，就需要比前一位小数所要求的项多十倍的项数。这样，计算时要有100位小数的准确度就需有2×10^{100}个项。我简直难于想像如何用手来加200个项，我是宁肯用计算机来算的；若还要计算更多的项，我肯定要坚持用计算机。现在让我用的计算机在一个部分和上再加一项要用10微秒左右（如果项很复杂花的时间更长）；我知道最先进的计算机可能还要快100倍，所以我们说是每微秒算10项，亦即每秒10^7项。这算很快了，但以这种速度算$\sum 1/n^2$的10^{12}项就需2×10^5秒，即约2.3天。没有哪个人愿意让我一个人使用计算机那么长的时间。如果我们很乐观地继续算到18位，我们就需用2×10^{10}秒，加起来已超过60年了。若要算到25位小数，大致需要宇宙的估计年龄那么长的时间，根本无法想象要算到100位时要用多少时间。所说这一切给我们的教训是，虽然有了无穷级数和的定义，凭一项一项地加的办法根本无法很准确地计算$\sum 1/n^2$。

至于别的级数，情况甚至更严重。要得到级数$\sum_{2}^{\infty} n^{-1}(\log n)^{2}$的两位小数位，要加$7\times10^{86}$个项；而$\sum_{3}^{\infty} n^{-1}(\log n)^{-1}(\log\log n)^{-2}$收敛得更慢，要得到两位小数的准确度，要加的项数我们在整个宇宙中也找不到足够量的物质让我们把它写上去（这是一个有10^{86}个数字符号的整数）。事实上在本书第三章中已经提醒过了，只有10^{80}个质点供我们在上面书写，我们只能在一个质点上写一个数字符号（我们事实上已经知道上面那个级数的和，精确到两位小数以上——大约是38.406768）。

你完全可以建议说，我们最好是等待更好的计算机出现。不过，有好些理由表明最小的有用的时间单位大致是10^{-23}秒（光线穿过一个电子所需的时间）；这个单位太小了，至今都还没有人给它起个名字。一架在10^{-23}秒钟便可以加一项的计算机，如果从宇宙诞生那一天开始计算，到现在也还没有加完10^{41}项哩！

对于教科书中的象$\sum f(n)$的任何一个收敛级数，尤其是关系到收敛的积分判别法时，都可以作类似的计算。第一个要求是要f减小到0；为了结果精确我们需要其几个导数都单调地趋于0的函数f。附表1中是一些数字结果（采自哈代〔4〕）。如果你想核实表中各个数字，或者自己想为另外的级数

造这类表，那么下面便是有关方面的几点评述．收敛得快的级数（从第5号级数起）用直接估计余项的办法就解决得很好了，比如把余项与几何级数的余项比较；用好的袖珍计算器、根据几何级数求和的公式就可得出幂级数的各条数据，这点已经讲过了．对于收敛很缓慢的级数，就得使用欧拉一麦克劳林公式（参看〔2〕），这时是将余项

$\sum_{n=N+1}^{\infty} f(n)$ 和 $\int_{N+1}^{\infty} f(t)dt$ 相比较。为此，

$\int_{x}^{\infty} f(t)dt$ 必须是列表函数；如若不是，可能就必须设法把它制成表。更准确些说，（对于“表现好”的函数）$\sum f(n)$ 的 N 项后的余项到头来几乎等于 $\int_{N+\frac{1}{2}}^{\infty} f(t)dt$。因此若要使余项小于 ε 就必须 $\int_{N+\frac{1}{2}}^{\infty} f(t)dt<\varepsilon$，或 $N+1/2<\psi(\varepsilon)$，此处 ψ 是函数 φ 的反函数，而 $\varphi(x)=\int_{x}^{\infty} f(t)dt$。例如对 $f(t)=t^{-3}$ 我们就有 $\varphi(x)=\frac{1}{2}x^{-2}$，因而 $\psi(y)=(2y)^{-\frac{1}{2}}$，但是事情并非总是这样简单。如果 φ 已经有很好的表了，原则上说来我们能够用逆插法得到 $\psi(y)$，但是在一些（看来）颇为简单的例子中我们却无法从已有的各表得到我们在表1中想得到的所有各项数据。（级数 $\sum 1/(n^2\log n)$ 正是一例。）

发散级数的情况也类似．这时没有余项，人们要问的自然就是：要使级数的部分和大于指定的数，比如10，100或10^6，我们必须加多少项？我们很自然地会求助于计算机——还有什么比让计算机去加到指定的和数更轻松呢？对某些级数来说，这很有效，而且事实上附表2的更小的各条数据就是这样计算出来的。（我们的数据有时比〔4〕中的数据更准确，原因是那些数据是60年前计算出的，当时没有计算机，没有袖珍计算器，甚至连象样的指数函数表也没有。）比如用计算机就容易发现，对于$\sum_{n=1}^{\infty}(n+1)^{-1}(\log(n+1))^{-1}$，要得到至少为3的和需要算8717项；但若要得到至少为4的和数我们就困难了，因为要计算5×10^{10}项。所以预先不晓得大致需要计算多少项便动手计算是很危险的。另外，若要加很长的部分和，我们应该从远端开始计算，因为远端的项小些；不然的话那些项计算了一会儿对计算来说好象都是零，于是各个部分和显然成了常数。

下面这段亲身经历的故事很清楚地说明了这一点．我曾经很想计算一个级数的和，这个级数事实上是收敛得很慢的级数（是"一些有名的整数序列"构成的稀疏了的调和级数），在另一个大学里的一位朋友对这个问题也很感兴趣，他请一位学生帮他

编了一个计算机程序来求和，把从第66430项起每隔$9^5=59049$（这是级数的结构决定的数）的间隔按八步算出的部分和数表寄了一份给我，有个什么东西使我想把表中的数区分开来，于是我发现第6与第7两数、第7与第8两数之差都等于0.05631、这时我方才看出计算机在每一步所加的都是同样的数$0.04631/59049=9.536\times10^{-7}$。我很想知道究竟这个数可能有什么意义，有一天半夜里突然醒来，发现这个数原来竟是十进制下2的一个负幂数。在2的幂数表中我发现$2^{-20}=9.537\times10^{-7}$。这就相当清楚了：我们的计算机计算到了20个二进数字，而且最后两段中的级数的项很小，到2^{-20}才归入。由于在这两段中级数的典型项在1.67×10^{-6}与1.25^{-6}之间，而$2^{-19}=1.9\times10^{-6}$，所以计算机也很可能不是把数目向最近的第20位二进数字纳入归整，而是舍去归整。

如果我们先进行一点理论推演工作，附表2中的大部分所列数目用好的袖珍计算器便可核实。我想简要指出这一点。当$f(x)$减小至零时，差

$$\sum_{n=1}^{N} f(n)-\int_{1}^{N} f(t)dt$$

趋于极限γ，对于$f(x)=\frac{1}{x}$（调和级数）这叫欧

拉常数．证明此点在微积分学中是个很容易的练习题（若绘出图来，做起来更令人信服）．于是推知部分和$\sum_{n=1}^{N} f(n)$（对于大N）必定趋近

$\int_1^N f(t)\,dt+\gamma$．若$\psi(y)$是$\int_1^y f(t)dt$的反函数，那么，要使$\sum_{n=1}^{N} f(n)$接近值A，必须要N几乎等于$\psi(A-\gamma)$．更细致的研究结果〔1〕表明，若写成$m=[\psi(A-\gamma)]$，则使得$\sum_{n=1}^{N} f(n)>A$的第一个指标N，即n_A，要么是m，要么是$m+1$，而且n_A几乎真的是最接近$\psi(A-\gamma)$的整数．A为整数值时，对于已经计算过的每种情形,都是真确的〔1，2,3〕．（所以附表2的大部分都可以在袖珍计算器上加以验证．）对于调和级数，$\psi(x)=e^x$，而n_A至少一直到$A=1000$都是最接近$e^{A-\gamma}$的整数〔5〕．一般地，对整值A来说，n_A是最接近$\psi(A-\gamma)$的整数，除非$\psi(A-\gamma)$碰巧比实际上看起来更接近一个奇数之半．所以当A大时（通常指比2大）要准确地求出n_A就需要求出$\psi(A-\gamma)$的小数后几位来．例如，对于调和级数，且$A=100$的情况我们就要让$e^{100-\gamma}$的一个值包括它整数部分的44位数字，并对小数部分

作出充分好的估计，保证它不要太接近一个整数加上$\frac{1}{2}$.幸好$e^{-\gamma}$是数论所关心的数，而J.M.乌伦奇又对精确计算有趣数字有爱好，所以当我对计算$e^{100-\gamma}$发生兴趣时，他才惠赐我$e^{-\gamma}$的150位近似值．至于e^{100}，那是计算机时代到来之前很久就由H.S.乌勒尔计算过了，比必要的精确度还计算得高。所以算是算得出来的，不过得乘它过够。

对大多数级数来说必须先计算γ。（说起来倒也有些奇怪，虽然原始欧拉常数很引人注目，其它的欧拉常数却有点被冷落了。）直接从定义出发来求不大容易，但若把欧拉一麦克劳林公式和计算机计算结合起来，常常就可以得到充分准确的值了.不过我们也需要把$\varphi(x)=\int_1^x f(t)dt$至低限度编成表；或者更好一些是把它的反函数$\psi$制成表。表2就是一些级数代表的有关数据。

很自然地，如果我们把两个发散级数$\sum f(n)$和$\sum g(n)$加以比较，并且知道$f(n)\to 0$比$g(n)\to 0$更快，我们就会以为$\sum f(n)$比$\sum g(n)$发散得更缓慢；换句话说就是，$\sum f(n)$的部分和要达到一个给定值所需的项数要比$\sum g(n)$所需的项数多些。事实上，发散得快的级数可能要花很长的

时间才赶得上来．一个富于启发性的例子便是$f(n)=n^{-0.1}$和$g(n)=1/\log(n+1)$，由于负幂数比对数的倒数更快地趋于零，所以$\Sigma f(n)$必定比$\Sigma g(n)$发散得更慢，而发散得快的级数要赶上来就得要有10^{14}项到10^{15}项才行。

参考文献

1. R.P.Boas, Jr., Growth of partial sums of divergent series, Math.of Comp., 31 (1977) 257—264.

2. —, Partial sums of infinite series, and how they grow, Amer.Math.Monthly, 84 (1977) 237—258,

3. —and J. W. Wrench, Jr., Partial sums fo the harmonic series, Amer. Math. Monthly, 78 (1971) 864—870,

4. G.H.Hardy, Orders of infinity, 2nd ed, Cambridge, 1924.

5. R.Spira, Personal communication.

九、开卜勒圆锥曲线

R.洪斯贝格尔

在三维坐标空间中，令z轴上的点光源T将以C（1,0,1）为心半经为1的球在xy平面上投下一个阴影（图46）。阴影的轮廓由从T射出并切于球的光线所成的圆锥确定，它就是这个正圆锥与xy平面的交线。因此按定义，包围阴影的曲线是一条圆锥曲线。

事实上从T射出的光线形成的锥体只是用来定义圆锥曲线的那个双叶锥体的一半．为了识别锥线阴影的类型，我们可以设想T的另一方有一个完全一样的锥体。得到的包围阴影的圆锥曲线的种类由T在z轴上的位置决定。若T在点S（0,0,2）之"上"，则xy平面只与圆锥的一半相交，相交的角变正好能产生出一个椭圆形的阴影来。T点恰好在S位置时xy平面与圆锥的母线平行，这时出现的

①这里记述的是(普林斯顿大学)A．M．塔克尔教授1970年10月26日在滑铁卢大学所作的通俗讲坐"开卜勒圆锥曲线"、我曾参加听讲，本文是根据我当时的笔记整理的。因此若有任何谬误，概不由塔克尔教授负责。

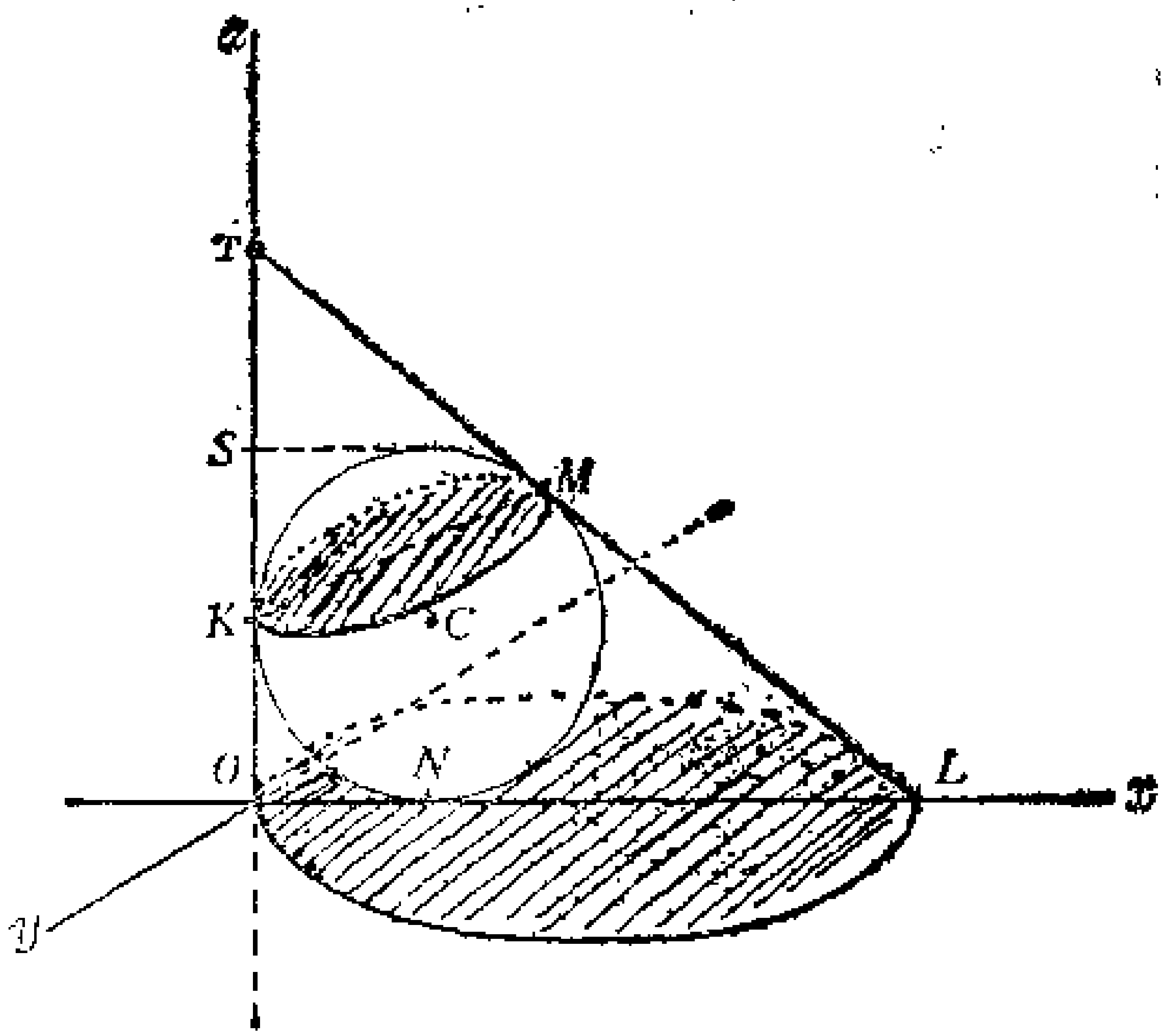


图46

阴影是抛物型的.而当T在S与点$K(0,0,1)$之间时，平面与圆锥的两叶都相交，产生一块双曲线阴影（虽然，如已经提到过的那样，只有一支实际上勾勒出了阴影的轮廓来），随着T点沿z轴向“上”移动，显然这阴影就趋近圆心为N $(1,0,0)$的单位圆.因此随着T点从K出发沿z轴向上移动，我们得到一族连绵不断变化的圆锥曲线，即从一系列双曲线变成抛物线，然后变成一系列椭圆.我们将会看到一切可以想象得出的圆锥曲线都包罗无遗了，因为对于离心率e的每个值，在这族曲线中都有一

条圆锥曲线相应地出现。（记住，除了圆以外，圆锥曲线可以定义为点的轨迹，这种点变动时与一个定点（焦点）的距离和它到一条定直线(准线)的距离之比是一个正常数（离心率）。）我们的目的就是要求这族圆锥曲线的方程；这项任务将把我们带到格外有趣的某些解析几何问题里去。

光线照亮了球的“冠”，其余部分依然在黑暗中。这明暗两个区域的交界线是一个圆，我好像记得听人家说过，登月宇航员把这个圆圈叫做“晨昏线”（terminator）。这不过就是切锥与球的切点组成的一个环而已。现在我们来看这个圈形含于xy平面中的横截面（见图46）。在这个横截面中，圆是从球那儿得来的，一对切线TM和TK是从圆锥那儿截出的，直径KM是“晨昏线”的。我们马上就会看到，在xz平面上，这条直径KM的斜率原来竟是在xy平面上勾勒出阴影轮廓的圆锥曲线的离心率！那也好，我们就预先用字母e来表示这个斜率吧。

在xz平面上K的坐标是$(0,1)$，因此KM在这平面上的方程就是

$$z-1=e(x-0),$$

也就是

$$1+ex-z=0.$$

在三维坐标系中，给定的球面方程为

$$(x-1)^2+(y-0)^2+(z-1)^2-1=0.$$

现在我们要做最奇怪的事情了。取球面方程左端和KM在xz平面上的方程的左端的平方，作出方程

$$(x-1)^2+y^2+(z-1)^2-1=(1+ex-z)^2.$$

乘出来并化简得到

$$y^2=2x(1+e-ez)+(e^2-1)x^2. \qquad (1)$$

然而这方程表示的是什么呢？原来它表示的是决定阴影形状的切锥的方程，证明这一点是满有趣味的。至此我们只消令方程中的$z=0$就得到圆锥阴影的方程。先试一下看，我们将得到的圆锥阴影的方程用e表示时是

$$y^2=2x(1+e)+(e^2-1)x^2.$$

详情以后再说。现在我们先来证明我们的奇怪方程实际上表示的是切锥。这是本文的主体。

首先我们引入变换

$$X=x, \qquad Y=y, \qquad Z=\frac{1+e}{e}-z.$$

这个变换的几何意义暂缓考虑，我们只来看看它对我们这个方程的作用。我们有 $eZ=1+e-ez$，于是我们造的方程（1）变成

$$Y^2=2X(eZ)+(e^2-1)X^2,$$

呈$pX^2+qY^2+rXZ=0$形。

不难看出这表示顶点在原点的一个圆锥。原点本身显然在这方程的轨迹上。假定另有一点 (a, b, c) 也在这轨迹上，这就会得出

$$pa^2+qb^2+rac=0.$$

于是把坐标 (ta, tb, tc)，t是任意实数，代进方程的左端，我们得到

$$p(ta)^2+q(tb)^2+r(ta)(tc)$$
$$=t^2\ (pa^2+qb^2+rac)\ =0,$$

这意味着 (ta, tb, tc) 也在轨迹上。但是当t取遍所有实数值时，点 (ta, tb, tc) 就要通过连接 (a, b, c) 与原点的直线上的每个点。也就是说连接 (a, b, c) 到原点的整条直线都在那个轨迹上点(ta, tb, tc)到原点的距离是点 (a, b, c) 离原点距离的t倍；参见图47）。

这轨迹与平行于XY平面（即Z为常数）的一个平面的交线方程如下

$$pX^2+qY^2+kX=0,$$

k是个适当的数值系数。但这个方程表示的是圆锥曲线，所以我们所探讨的轨迹就是这条圆锥曲线与作为顶点的原点所产生的圆锥。讨论到这一步我们并不知道这是一个正圆锥，但这不是一个问题。

我们已经知道，变换后的方程表示一个圆锥。但是我们开初的那个方程又表示的什么呢？现在就

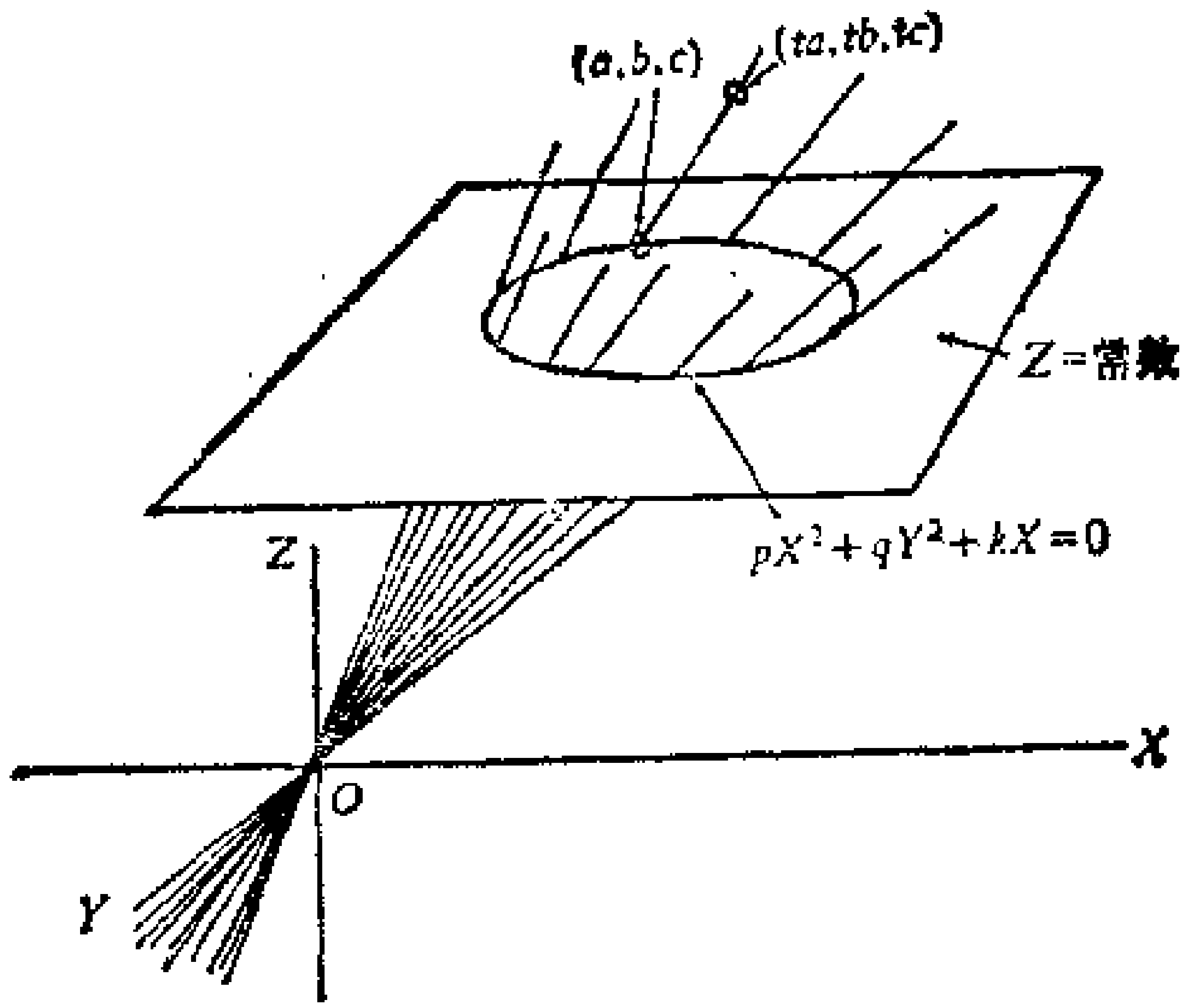


图47

该来分析我们的变换蕴涵什么几何意义了。我们曾用同样的x坐标（$X=x$），同样的y坐标，至于z坐标呢，我们用的是$-z$再加上一个（$1+e$）/e。用$-z$不过是从xy平面把一切加以反射而已，加上（$1+e$）/e则表示作了一个平移。所以总效果就是经xy平面的一个反射和在正z轴的方向上作大小为（$1+e$）/e的一次平移。而这样的变换不会使形状和长度遭到扭曲。要是一个轨迹的最终图像是一个圆锥的话，那么它原先必定已经就是圆锥了。所以我们的原始方程（1）的的确确代表一个圆锥。

接着我们要做的工作就是证明它是切线光线形成的圆锥。

我们来消除上面的变换：把作为像的圆锥在z轴方向上平移一个距离$-(1+e)/e$，然后经xy作一次反射。这样做就会先把顶点（0，0，0）往下移至$(0,\ 0,\ -(1+e)/e)$，然后又往上移至点$(0,\ 0,\ (1+e)/e)$，于是得知初始圆锥的顶点是$(0,\ 0,\ (1+e)/e)$。

现在来考虑晨昏线上的任意点$P(x,\ y,\ z)$，由于光源T在z轴上，晨昏线所在平面与yz平面斜交成一定角度，于是相交于一条直线v，此线与y轴平行（图48）。所以过点P平行于y轴的一条直

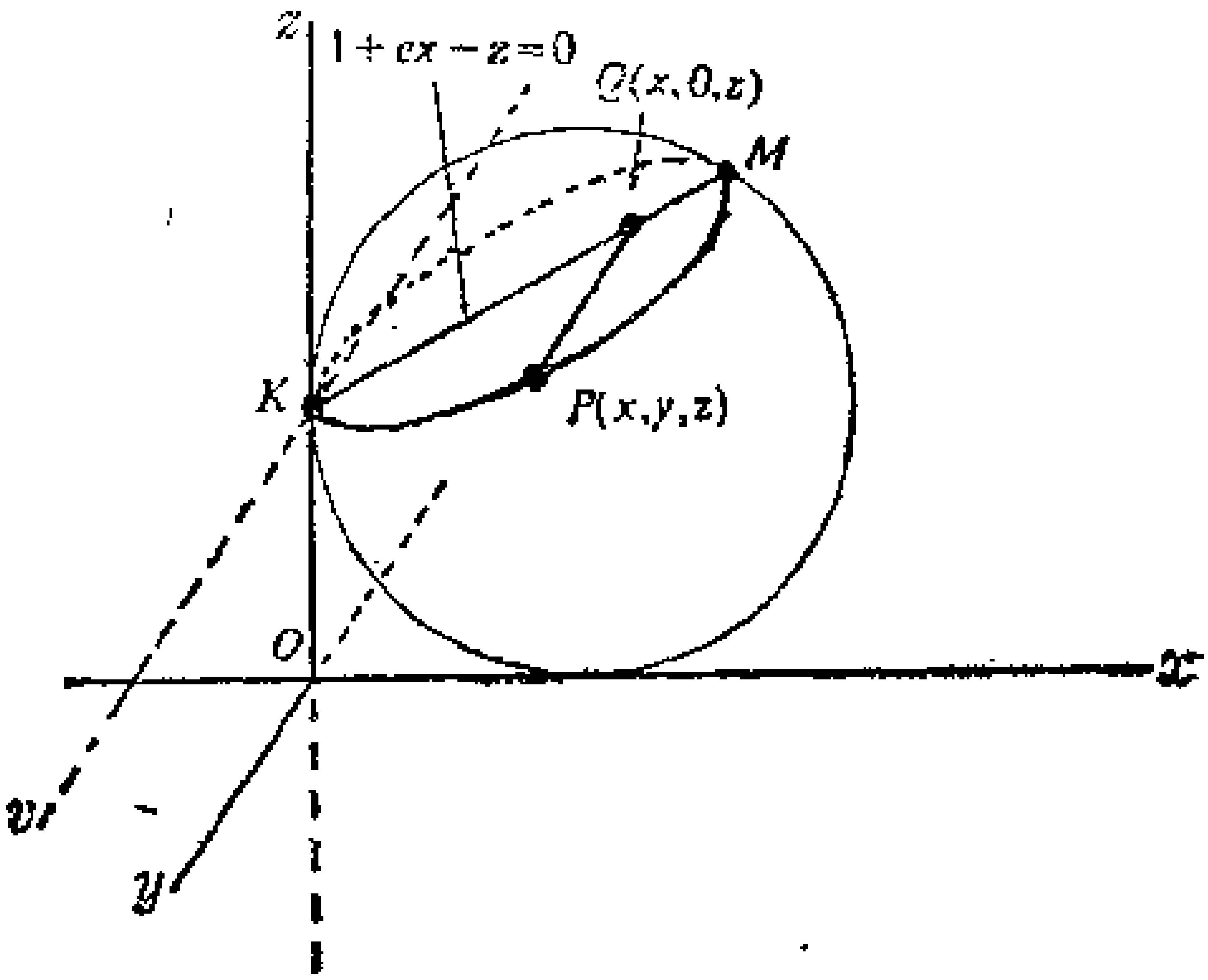


图48

线就会穿越晨昏线所在的平面而与直径KM相交于一点$Q(x, 0, z)$，此点除y坐标为零外其它坐标与P相同，因为直径在xz平面上。因为Q在直径上，所以它的坐标满足直径的方程，即

$$1+ex-z=0.$$

但是这x与z和P的坐标一样。由于y不出现在方程中，所以即使是P的坐标也要使$1+ex-z$取0值。然而P是球面上的一个点，所以它的坐标满足方程

$$(x-1)^2+y^2+(z-1)^2=0$$

因此点P的坐标满足我们想出来的方程（1），因为我们刚才看到了它使得方程两边取零值。但P是晨昏线上的任意点，因此整条晨昏线必定属于所说的那个轨迹。所以这个轨迹必然是晨昏线和顶点$(0, 0, (1+e)/e)$所产生的圆锥。

现在我们看到虽然光源T的位置在z轴上是不受限制的（只要它在K上方），然而只有点T的一个位置才会在xz平面上使KM有一个特别的斜率e。所以我们就可以反过来从一个具体的e值倒推出相应的光源T的坐标。我们的切线锥之所以特别，是因为它的顶点就是切球面于M点的平面横截z轴的那个T点。这一切说明在xz平面上的横截面上，直线TM是球的圆截部分KMN的一条切线（见图49）。我们想证明在3维空间中T是$(0, 0, (1+e)/e)$或者在xz平面上就是$(0, (1+e)/e)$为此目的我

们来求圆KMN的方程，M的坐标，切线TM的方程以及最后求出T的坐标。

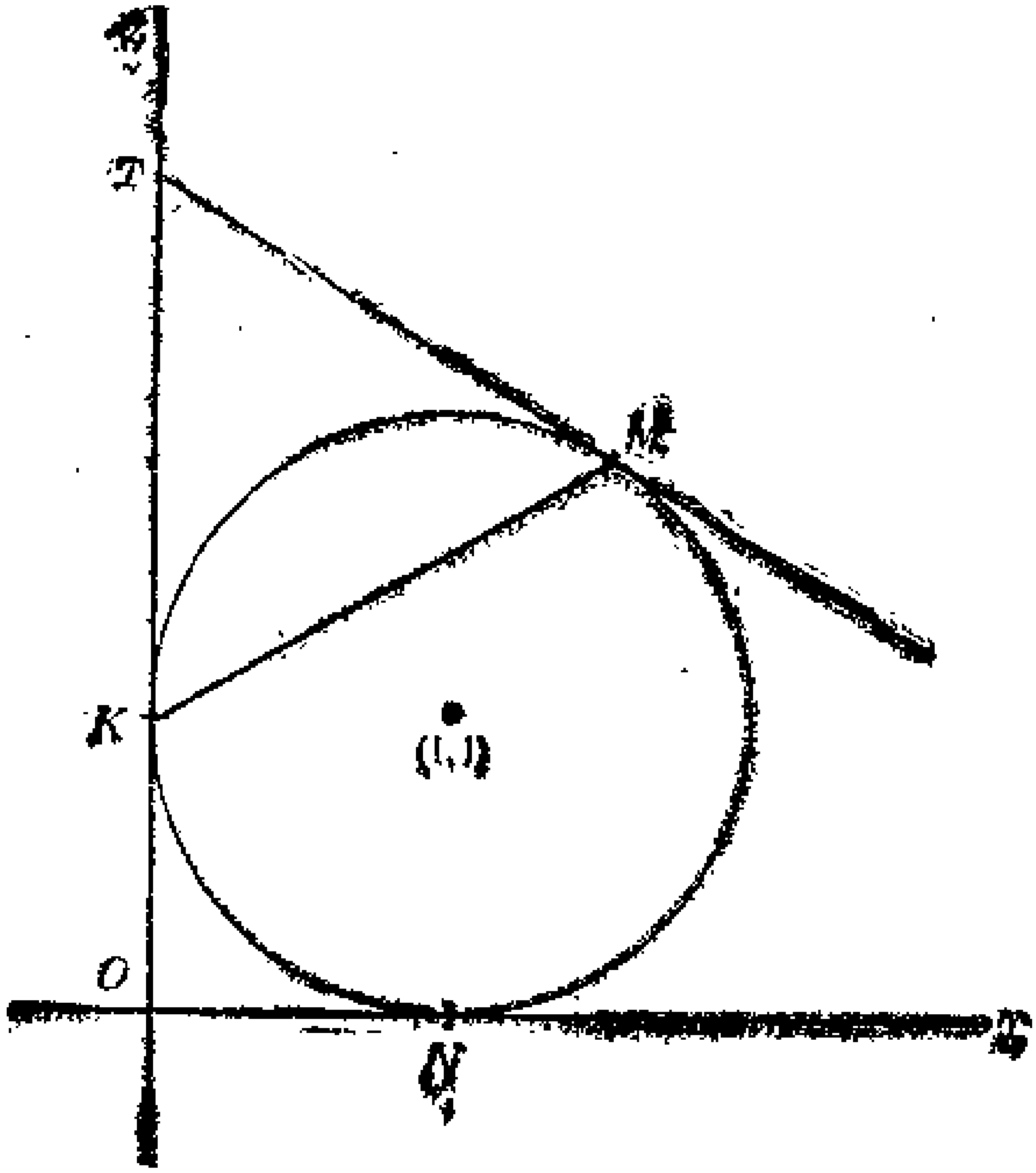


图49

圆KMN的中心为（1，1），半径为1，所以方程为

$$(x-1)^2+(z-1)^2=1$$

我们早先就知道KM的方程是$z-1=ex$。解之得

$$(x-1)^2+e^2x^2=1,$$

即

$$(1+e^2)x^2-2x=0,$$

其根$x=0$　给出点K，另一根

$$x=\frac{2}{1+e^2}$$

属于M。于是对M来说有

$$z=1+ex=1+\frac{2e}{1+e^2}=\frac{(1+e)^2}{1+e^2}.$$

所以M就是点

$$\left(\frac{2}{1+e^2},\ \frac{(1+e)^2}{1+e^2}\right)$$

圆的方程很容易就化简成

$$x^2+z^2-2x-2z+1=0,$$

由于在点(x_1, z_1)切于圆$x^2+z^2+2gx+2fz+h=0$的切线的方程是$x_1x+z_1z+g(x+x_1)+f(z+z_1)+h=0$，所以过$M$点的切线方程是

$$\frac{2}{1+e^2}x+\frac{(1+e)^2}{1+e^2}z-\left(x+\frac{2}{1+e^2}\right)$$

$$-\left(z+\frac{(1+e)^2}{1+e^2}\right)+1=0$$

对于我们一心想求的z截距，我们令$x=0$ 得

$$\frac{(1+e)^2}{1+e^2}z-\frac{2}{1+e^2}-z-\frac{(1+e)^2}{1+e^2}+1=0,$$

$(1+e)^2z-2-(1+e^2)z-(1+e)^2+1+e^2=0$,

$$2ez-2-2e=0$$

亦即

$$z=\frac{1+e}{e},$$

这就是要求的。

因此我们所谈的圆锥的确表示的是那个切线锥，而阴影的方程则是

$$y^2=2x(1+e)+(e^2-1)x^2.$$

重新排列得$x^2-2x+y^2=e^2x^2+2ex$，两边加1得

$$(x-1)^2+y^2=(ex+1)^2,$$

亦即

$$(x-1)^2+y^2=e^2\left(x+\frac{1}{e}\right)^2.$$

因而

$$\sqrt{(x-1)^2+y^2}=e\left(x+\frac{1}{e}\right)$$

这个式子的意思就是：点(x,y)到$(1,0)$的距离是它到直线$x+(1/e)=0$的距离的e倍．也就是说，(x,y)到$(1,0)$的距离与到直线$x+(1/e)=0$的距离之比是一个正常数e．所以我们的轨迹是焦点为$N(1,0)$，准线为$x=-(1/e)$，离心率为e的一条圆锥曲线．我们观察到，每一正实数e都给出一个在$K(0,0,1)$之"上"的顶点$T(0,0,(1+e)/e)$从而得出我们圆锥曲线族中一切可能形状的曲线，但圆不在此例，因为我们已经看到，圆可看成是T沿着正z轴向上移动的极限情况．

最后我们注意到，圆锥阴影方程的形式是

$$y^2=px+qx^2,$$

（式中$p \neq 0$），伟大的希腊几何学家阿波龙尼乌斯在钻研圆锥曲线时就得到了圆锥曲线的一般方程，他把圆锥曲线分类如下（在我们的情形$q = e^2 - 1$）：

$q < 0$，有$y^2 < px$：椭圆（注意这就得出上面的$e < 1$），

$q = 0$，有$y^2 = px$：抛物线（$e = 1$），

$q > 0$，有$y^2 > px$：双曲线（$e > 1$）。

抛物线只有一个焦点曾使开卜勒很困恼。然而在上面的曲线族中，抛物线在椭圆与双曲线之间，后两种曲线都有两个焦点。这使他非常高兴，因为他看出快成抛物线时椭圆的一个焦点趋向无穷。于是他说抛物线有两个焦点，但是其中一个是“盲点”，即位于无穷远处的点。

开卜勒首次使用“焦点”这个术语，虽然更早得多的时候人们就知道它的重要性了。他是第一个研究圆锥曲线连续族的人，也是第一个指出一个焦点在无穷远处的人。

本文结束时我们想指出：我们讨论的球面现在称为“丹德兰球面”，是纪念十九世纪二十年代的一位比利时数学家；我们的圆锥曲线族中每条曲线都以点N（1，0，0）为一个焦点，以原点为顶点。

参考文献

1. J. Kepler, *Ad Vitellionem Paralipomena*, 1604.
2. W. M. Ivins, Jr., Art and Geometry, Dover, New York.

十、斯基威斯数

R. P. 波亚士

斯基威斯数（Skewes这个名字的读音我觉得应读成两个音节）在大约四十年前在数学界是人人皆知的数，但是看来对这一代人就已经是陌生的东西了。这个数是G.H.哈代宣传开的；在〔6，*p.* 17〕中他写道"我想，这是数学中有确定作用的最大的数了。"虽然最近还有更大的数具有确定性的意义（参见〔2〕，特别是〔23〕），不过斯基威斯数仍然值得拿来思考。它是什么数，是什么意义，最近情况如何，我很想加以介绍。(到1953年为止的一份十分专门的讨论收在李特尔武德的著作〔16〕中；而一份通俗讨论——几乎不用什么数学！则是由伊萨克·阿西莫夫〔1〕给出的。)

我们需要知道一些背景知识。早在两千多年前希腊人就在研究素数了（欧几里德曾证明素数有无穷多个）；但就我所知，希腊人从未想过数一数到一个给定值x为止有多少素数。事实上好象是没有人想到过这样做，一直到1800年前不久勒让德〔9〕和高斯〔4〕才考虑$\pi(x)$，即小于或等于x

的素数的个数；他们是就当时已有的数据在一些区间上考虑的.当时（1796年）素数表刚编到400313.勒让德用启发性的推理证明素数密度，即$\pi(x)/x$，对于大的x应近似等于$1/(\log x-B)$，式中的B是他根据经验确定的常数。在素数序列中发现（至少是显得来是发现）了如此大的规整性，在当时简直引起了轰动。高斯走的方向稍有不同。他根据数值性数据猜测在x处的"单位区间素数的平均个数"大约是$1/\log x$，因此他正象现在任何一位学过微积分的学生一样推断，$\pi(x)$应近似地等于

$$\int_2^x \frac{dt}{\log t},$$

到1849年时素数表更长了，高斯的公式看起来就更准确了；勒让德公式也是这样。

根据有限的数据凭经验作出的估计不消说是可能出错的。以1，2，4，8，16作前几项的数列下一项是什么？照科学（而不是数学）归纳法，人们多半会回答说是"32"。然而我们大家都知道，逻辑正确的回答是，"我不知道。"的确，把专门搞的构造除开，下一个数完全有理由是21，22，31或36（参见〔21〕，序列425—429，433）。

不过从某个意义来说，勒让德和高斯都猜对了；他们两人很可能都想使我们相信，这个近似公

式是渐近地正确的公式，换句话说当$x\to\infty$时

$$\pi(x)\Big/\frac{x}{\log x-B}\to 1\quad\text{或}$$

$$\pi(x)\Big/\int_2^x\frac{dt}{\log t}\to 1.$$

这两个极限式实际上是等价的（想了解必要的技巧，请参见〔5〕），都等价于

$$\pi(x)\Big/\frac{x}{\log x}\to 1.$$

这最后的式子就是“素数定理”；虽然在1800年前后实际上就已经猜测到了，但是直到1896年才获得证明（阿达马与德·拉·瓦莱·布森分别独立地证明的）。

数学家是永远不满足的；一切定理都可能再搞精确些。在这个例子中接着要问的问题是“误差有多大？”我们知道$\pi(x)\Big/\frac{x}{\log x}\to 1$，但是，即使$\pi(x)-\frac{x}{\log x}\to\infty$时，这也是成立的。德·拉·瓦莱·布森证明，高斯的猜测优于勒让德的猜测；不管给B什么值，高斯公式的误差始终要小些。所以人们喜欢把$\pi(x)$与$\int_2^x\frac{dt}{\log t}$相比，或者更经常的是，把$\pi(x)$与更对称的积分$\mathrm{li}\,x=\int_0^x\frac{dt}{\log t}$相比。在$t=1$处这是个反常积分，因为被积函数此时等于零；这积分应理解为主值，当$x>1$时的

$$\mathrm{li}\,x=\lim_{h\to 0}\left(\int_{0}^{1-h}\frac{dt}{\log t}+\int_{1+h}^{x}\frac{dt}{\log t}\right)$$

（在$x=1$处，$\mathrm{li}\,x$无定义，不过我们也只对大x感兴趣。）

施行变量替换$\log t=u$，$\mathrm{li}\,x$化为指数积分

$$\mathrm{li}\,x=\int_{-\infty}^{\log x}u^{-1}e^{u}\,du,$$

比起$\mathrm{li}\,x$来，这个积分编成的表更全，计算起来也更容易。

随着关于各个素数的分布情况的知识愈来愈丰富，也就可能对越来越多的x值将$\pi(x)$和$\mathrm{li}\,x$加以比较。即使在今天，超过10^7的一个个的素数也知道得不多，但是不必一个个地了解就求出小于给定数的素数究竟有多少的方法却是有的（参见〔11〕和〔13〕）。下面的表是从〔7〕，〔11〕，〔13〕摘来的，$\mathrm{li}\,x$的值取为最接近的整数。

x	$\pi(x)$	$\mathrm{li}\,x$	$\mathrm{li}\,x-\pi(x)$	$\pi(x)/\mathrm{li}\,x$
10^3	168	178	10	0.9438
10^4	1229	1246	17	0.9861
10^5	9592	9630	38	0.9961
10^6	78498	78628	130	0.9983
10^7	664579	664918	339	0.9995
2×10^7	1270607	1270905	298	0.99977
3×10^7	1857859	1858213	354	0.99981

3.3×10^7	2031667	2031974	307	0.99985
3.7×10^7	2261623	2262253	630	0.99972
4×10^7	2433654	2434016	362	0.99985
10^8	5761455	5762209	754	0.99987
1.1×10^8	6303309	6303656	347	0.999945
1.2×10^8	6841648	6842446	798	0.999883
1.6×10^8	8974458	8975397	939	0.999895
1.8×10^8	10030385	10030768	383	0.999962
5×10^8	26355867	26356832	965	0.999963
7.5×10^8	38703181	38705046	1865	0.999952
10^9	50847534	50849235	1701	0.999967
10^{10}	455052512	455055615	3101	0.999993

10^7与10^9之间的各条数据x选自已出版的一些表，目的是让大家看看差$\mathrm{li}x-\pi(x)$的摆动有些还是比较大的。注意，$\pi(x)/\mathrm{li}x$看来的确是趋近1，这符合素数定理，虽然它并不是单调地趋近于1。然而$\mathrm{li}x-\pi(x)$始终是正的，而且总的说来随着x增大似乎还在增大。我们认为$\mathrm{li}x-\pi(x)\to+\infty$，这倒是情有可原的。1914年以前人们猜测$\mathrm{li}x-\pi(x)$至少总是正数，不仅数值证据提示出这一点，而且存在一个无穷级数表示$\mathrm{li}x-\pi(x)$这级数从$+\frac{1}{2}\mathrm{li}(x^{\frac{1}{2}})$开始，给出的数值结果远比$\mathrm{li}x$本身好得多，这个证据看来绝对有利于说$\mathrm{li}x-\pi(x)$是正数。但是李特尔伍德1914年却证明$\mathrm{li}x-\pi(x)$在无穷多次的变正负号（[15]，[7]

中有更易懂的证明）。原来$\mathrm{li}\,x-\pi(x)$的公式中的正项最终被还多含了因子$\log\log\log x$的负项抵消完了，这个因子的影响要在很长时间以后才显现出来（$\log\log\log 10^{20}$比1.343还小）。李特尔伍德的证明是纯粹存在性证明的经典范例；它连使$\pi(x)>\mathrm{li}\,x$的x的一个例子也没举出，甚至连X的例子也没举出过，对这种X当$x<X$时有$\pi(x)>\mathrm{li}\,x$。

大约是在1930年，李特尔伍德的学生斯基威斯开始寻求一种证明，这个证明可以给出一个特别的X使得x达到X时$\mathrm{li}\,x-\pi(x)$改变符号。到1933年他终于找到了这样的X，当时他的结果〔19〕是

$$X=10^{10^{10^{34}}},$$

前题是假定黎曼假设成立。（我隔一会儿再来解释黎曼假设。）正是这个X以斯基威斯数而著称于世，它确实是个大数。为了对它有好大有个概念，不妨回忆一下卡斯纳的谷戈勒克数*〔8，*p*. 23〕它只不过是：

$$10^{10^{10^{2}}},$$

---

*）数学家卡斯纳（E. Kasner，1878—1955）把10^{100}叫googol，把$10^{10^{100}}$叫googolplex——译者注。

而宇宙中的质点个数只不过大约是$10^{80}<10^{10^2}$〔14，p.1216〕。哈代以前常常说〔6，p. 17〕，斯基威斯数有点象是用这些质点来下棋时的局数，这里的一步棋是指交换两个质点，同一位置第三次出现则一局棋算结束。(李特尔伍德〔16，p. 109〕证明了，普通棋可能的局数小于$10^{10^{71}}$。)

再开始讲最近的情况之前，我必须把黎曼假设解释一下（更详细的介绍请读〔3〕，〔22〕，〔25〕）。黎曼ζ函数当s的实部大于1时由公式

$$\zeta(s)=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\cdots$$

定义。此级数对其它的s值是发散的。但是$\zeta(s)$可以开拓到（除点$s=1$外的）整个复平面，这和人们更熟悉的Γ函数可以开拓的情况有点类似。开拓后的函数当s为负整数时取值为零，在带形$0<\mathrm{Re}\,s<1$之外别无其它零点。在此带形之内，$\zeta(s)$有无穷多个零点，实际上对于$|\mathrm{Im}\,s|<T$比$T\log T$的一个常数倍数还多。黎曼假设说，除了在负整数处的零点（叫"平凡零点"）外，$\zeta(s)$的全部零点的实部都是$\frac{1}{2}$。虽然现在我们比黎曼对ζ函数懂得更多得多，但是黎曼假设至今仍未得到证明。实部为$\frac{1}{2}$的零点有无穷多个，最近勒维逊〔12〕指出，三分之一以上的零点实部是$\frac{1}{2}$。罗赛尔、纯菲尔德和他

们的合作者们计算了头三百五十万个零点的数值（关于详情，请读〔18〕）；R·P·勃伦特〔24〕宣布他计算了4千万个零点。计算了这么多，它们的实部都是$\frac{1}{2}$。

ζ函数与素数的联系在无穷乘积公式（对$s>1$成立）

$$\zeta(s)=\prod\frac{1}{1-p^{-s}},$$

在此式中（和所有类似公式中）都假定p取遍相邻素数。这并没有马上指明$\zeta(s)$的零点如何用于素数分布的理论。不过从这个公式我们却可以得到

$$\log\zeta(S)=\sum_{p,m}\frac{1}{mp^{ms}},$$

（这里的m取遍正整数，p取遍素数），于是

$$-\frac{\zeta'(s)}{\zeta(s)}=\sum_{p,m}\frac{\log p}{p^{ms}}.$$

后两个公式都肯定地意味着$\zeta(s)=0$的那些点一定会和素数分布有某种关系。但精确的关系却很难看出来。我们想把一个公式倒过来，用ζ函数来表示素数，而不是让素数去表示ζ函数。首先，求出用ζ函数表达的$\psi_1(x)$而不是求$\pi(x)$要容易得多，这里$\psi_1(x)=\int_0^x\psi(u)\,du$而$\psi(x)=$

$\sum_{p^m \leq x} \log p$. 然后再来求$\pi(x)$ 本身，这时只是技术环节上麻烦些而已，而不是真的很困难.真正的问题是，甚至$\psi_1(x)$ 也是由复平面上的包含了$\zeta'(s)/\zeta(s)$的一个积分来表达.这里，关于ζ函数零点的位置我们知道得越多，我们获得的情况就越多；所以黎曼假设与素数分布有关系。当然，即令黎曼假设成立，它也并不告诉我们有关ζ函数零点的全部情况：它只是把这些零点放到了一条线上，却并没有说明这些零点究竟在线上的何处。然而我们却可以说ζ函数零点位置问题是和素数分布问题等价的。与ζ函数零点有关的一个物理学问题颇为有趣。这是分析黑体混合物发出的辐射从而确定黑体间的温度分布的问题。这个问题可以表达成一个积分方程，要解这个方程必需用ζ函数去除某个东西〔17，p.41〕.所以维纳以前常常说分析黑体辐射的物理问题和确定素数分布的纯算术问题是等价的。

如果黎曼假设成立，就可以比（至少在目前）它不成立时证明出关于素数分布的更精确的定理来。例如，我们有$\pi(x)/\mathrm{li}\,x \to 1$趋近得有好快？我们有$\pi(x)/\mathrm{li}\,x = 1 - \varepsilon(x)$，此处$\varepsilon(x) \to 0$。如果没有黎曼假设，我们知道的最佳结果是

$$\varepsilon(x) < \exp\left(-c(\log x)^{\frac{1}{2}}\right),$$

式中c是正常数；如果有黎曼假设，上述结果就可

以致善为

$$\varepsilon(x)<cx^{-\frac{1}{2}}(\log x)^2=c\exp(-\tfrac{1}{2}\log x+2\log\log x)$$

（c与前面不同），这是更强得多的结论。

现在我们回头来看斯基威斯数。在证明存在定理时，李特尔伍德可以用黎曼假设，因为如果黎曼假设为谬则很容地就得到存在定理。然而斯基威斯却发现，不要黎曼假设求显式估值比要了黎曼假设困难些；1937年他用

$$10^{10^{10^{10^{20}}}}$$

取代了最初的斯基威斯数（没有用未证明的假设条件）。这个新数当然更是大得不可想象。（1955年他把最顶上的指数29变小为3。）实际上斯基威斯并没有用黎曼假设及其否定形式而是用的黎曼假设的一项推论及其否定形式；想知道详细的讲解，请读〔16〕的113—115页。总之，如果人们可以说已经知道了什么取决于很复杂的证明的数学事实的话，那么在1955年就已经知道了，对于小于上述特别数的某些x，$\mathrm{li}\,x-\pi(x)$要变号。

你也许觉得很纳闷，为什么没有哪个人把这个问题拿到计算机上去解决。没有这样做的一个原因是，直接计算似乎办不到。看来很可能，使$\pi(x)$

$>\mathrm{li}\,x$的第一个数x已经比10^{20}大了，而这个数是大大超出计算机的能力范围（人们认为至今的宇宙史才经历了不到10^{24}秒〔14，*p*．797〕．不过却可以借助计算机去间接地攻这个问题．莱曼〔10〕不直接去对付$\pi(x)-\mathrm{li}\,x$而是先作尺度变换$x=e^u$，把那些数变小到可以对付的范围内，于是得到显明公式，这公式不是表示$\pi(x)-\mathrm{li}\,x$本身的而是表示一个平均值$ue^{-u^{1/2}}\left\{\pi(e^u)-\mathrm{li}(e^u)\right\}$．然后他就证明了斯基威斯数可换成$10^{10^{3.067}}<2\times 10^{1165}$．因为是取的平均，他未给出使得$\pi(x)>\mathrm{li}\,x$的具体的$x$，但他的确证明了在$1.53\times 10^{1165}$和$1.65\times 10^{1165}$之间这种数的个数比$10^{500}$还大，所以哈代对有人会找到使$\pi(x)>\mathrm{li}\,x$的显式$x$所持的怀疑〔6，*p*．17〕，可能在不太远的将来被消除．

参考文献

1. I. Asimov, Skewered! , Fantasy and Science Fiction, November, 1974, 131—141.

2. A. Baker and J. Coates, Integer Points on curves of genus 1, Proc. Cambribge Philos. Soc., 67(1970)595—602.

3. H. M. Edwards, Riemann's Zeta Functi

ion, Academic Press, New York and London, 1974.

4. C. F. Gauss, Tafel der Frequenz der Primzahlen; Letter to Encke. Werke, vol, 2, 2nd ed, Gottingen, 1876, 436—443, 444—447.

5. G. H. Hardy, Orders of Infinity: The 'Infinitarcalcul' of Paul du Bois-Reymond. 2nd ed, Cambridge University Press, 1924.

6. ——, Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, Cambridge University Press, 1940.

7. A. E. Ingham, The Distribution of Prime Numbers, Cambridge Univesity Press, 1932.

8. E. Kasner and J. Newman, Mathematics and the Imagination, Simon and Schuster, New York, 1940.

9. A. M. Legendre, Essai sur la theorie des nombrse, 2nd ed, Courcier, Faris, 1808, 394—398.

10. R. S. Lehman, On the difference π(x)—lix, Acta Arith, 11 (1966) 397—410.

11. D. H. Lehman, On the exact number of primes less than a given limit, Ill-

inois J. Math.,3 (1959) 381—388.

12. N. Levinson, More than one third of zeros of Riemann's zeta-function are on $\sigma=1/2$, Advances in Math., 13(1974)383—436.

13. D. C. Mapes, Fast method for computing the number of primes less than a given limit, Math. of Comp, 17 (1963) 179—185.

14. C. W. Misner, K. S. Thorne, and J. A. Wheeler, Gravitation, Freeman, San Francisco, 1973.

15. J. E. Littlewood, Sur la distribution des nombres premiers, C. R. Acad. Sci. Paris, 158 (1914) 1869－1872.

16. ———, A Mathematician's Miscellany, Methuen, London, 1953.

17. R. E. A. C. Paley and N. Wiener, Fourier Transforms in the Complex Domain, Amer. Math. Soc., New York, 1934.

18. J. B. Rosser and Lowell Schoenfeld, Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$, Math. of Comp, 29 (1975) 243－269

19. S. Skewes, On the difference $\pi(x)$

$-\mathrm{li}\,x$, J. London Math. Soc., 8 (1933) 277–283.

20. ——, On the difference $\pi(x)-\mathrm{li}\,x$ II, Proc. London Math. Soc, (3) 5 (1955) 48–70.

21. N. J. A. Sloane, A Handbook of Integer Sequences, Academic Press, New York and London, 1973.

22. E. C. Titchmarsh, The Theory of the Riemann Zeta-Function, Clarendon Press, Oxford, 1951.

23. M. Gardner, Mathematical games, Scientific American (Nov. 1977) 28.

24. R.P. Brent, The first 40,000,000 zeros of $\xi(s)$ lie on the critical line, Notices Amer. Math. Soc., 24 (1977) A417.

25. H. L. Montgomery, Zeta zeros on the critical line, Amer. Math. Monthly, 86 (1979) 43–45.

［Ｇｅｎｅｒａｌ　Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝数学小品译丛　数学瑰宝（第三辑）
作者＝
页数＝２４２
ＳＳ号＝０
出版日期＝